

# *Urban morphology parameters from global digital elevation models: implications for aerodynamic roughness for wind-speed estimation*

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Kent, C. W., Grimmond, S., Gatey, D. and Hirano, K. (2019) Urban morphology parameters from global digital elevation models: implications for aerodynamic roughness for wind-speed estimation. *Remote Sensing of Environment*, 221. pp. 316-339. ISSN 0034-4257 doi: <https://doi.org/10.1016/j.rse.2018.09.024> Available at <https://centaur.reading.ac.uk/79656/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.rse.2018.09.024>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Urban morphology parameters from global digital elevation models: implications for aerodynamic roughness for wind-speed estimation

Christoph W. Kent<sup>1\*</sup> • Sue Grimmond<sup>1\*\*</sup> • David Gatey<sup>2</sup> • Kohin Hirano<sup>3</sup>

<sup>1</sup>Meteorology, Univ. of Reading, Reading, UK <sup>2</sup>RMS, London, UK <sup>3</sup>National Research Instit. for Earth Science & Disaster Resilience, Ibaraki, Japan

\*E-mail: [C.W.Kent@pgr.reading.ac.uk](mailto:C.W.Kent@pgr.reading.ac.uk) \*\*E-mail: [c.s.grimmond@reading.ac.uk](mailto:c.s.grimmond@reading.ac.uk)

- *ASTER, SRTM and TANDEM-X elevation data compared to benchmark data in London, UK*
- *TanDEM-X is most like benchmark data (pixel-to-pixel, geometric and roughness parameters)*
- *Empirical corrections to TanDEM-X parameters developed from five global cities*
- *Corrections improve estimation of parameters and for a separate urban location*
- *Wind-speed estimates with elevation model parameters compared to observations*

## Abstract

Urban morphology and aerodynamic roughness parameters are derived from three global digital elevation models (GDEM): Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Shuttle Radar Topography Mission (SRTM), and TanDEM-X. Initially, each is compared to benchmark elevation data in London (UK). A moving window extracts ground heights from the GDEMs, generating terrain models with root-mean-square accuracy of up to 3 m. Subtraction of extracted ground heights provides roughness-element heights only, allowing for calculation of morphology parameters. The parameters are calculated for eight directional sectors of 1 km grid-squares. Apparent merging of roughness elements in all GDEMs causes height-based parameter underestimation, whilst plan and frontal areas are over- and under-estimated, respectively. Combined, these lead to an underestimation of morphometrically-derived aerodynamic roughness parameters. Parameter errors are least for the TanDEM-X data. Further comparison in five cities (Auckland, Greater London, New York, Sao Paulo, Tokyo) provides basis for empirical corrections to TanDEM-X-derived geometric parameters. These reduce the error in parameters across the cities and for a separate location. Meteorological observations in central London give insight to wind-speed estimation accuracy using roughness parameters from the different elevation databases. The proposed corrections to TanDEM-X parameters lead to improved wind-speed estimates, which combined with the improved spatial representation of parameters across cities demonstrates their potential for use in future studies.

## Keywords

• Aerodynamic roughness length • Digital elevation model • Digital surface model • Digital terrain model • Morphometric roughness methods • Urban • Wind speed profile • Zero-plane displacement

## 1 Introduction

Accurately resolving urban morphology is central to many research areas, including: urban planning, human settlement and land use mapping, change detection (e.g. urban sprawl) and the monitoring/ management of risk and disaster. The morphology is also critical for modelling and understanding the urban climate (Grimmond and Souch 1994, Yan et al. 2015). The type, presence and distribution of surface roughness elements (urban morphology) influences the storage and turbulent fluxes of heat and other scalar quantities such as pollutants (Gál and Unger 2009, Grimmond et al. 2010, 2011, Lindberg and Grimmond 2011b, Ng et al. 2011, Garuma 2017). Accurate representation of urban morphology is also vital to accurately model fluxes of momentum. Surface roughness elements influence the spatially- and temporally-averaged properties of the air flow, as well as the turbulent characteristics. In combination, this helps define the structure of the urban boundary layer (Roth 2000, Martilli 2002, Arnfield 2003, Britter and Hanna 2003, Fernando 2010).

The influence of the surface upon momentum fluxes can be characterised using aerodynamic roughness parameters (zero-plane displacement,  $z_d$ , and roughness length,  $z_0$ ). With these parameters, the spatially- and temporally-averaged wind-speed profile can be estimated if the flow is free from roughness-element wakes. The extent of vertical displacement of the wind-speed profile is indicated by  $z_d$ , which may correspond to the ‘drag centroid’ or height the mean drag appears to act (Thom 1971, Jackson 1981). The  $z_0$  is the height wind speed becomes zero in the logarithmic wind-speed profile, in the absence of  $z_d$  (Blackadar and Tennekes 1968). The morphology of roughness elements can be used to determine  $z_d$  and  $z_0$  using morphometric methods (Grimmond and Oke 1999, Kent et al. 2017a). Therefore, elevation databases that resolve roughness-element morphology allow estimation of  $z_d$  and  $z_0$  in cities and the associated: wind speeds (e.g. Millward-Hopkins et al. 2012, 2013, Varquez et al. 2015, Kent et al. 2017a, 2018), surface heat fluxes (e.g. Crawford et al. 2018) and surface shear stress in meso-scale meteorological models (e.g. Weekes and Tomlin 2013).

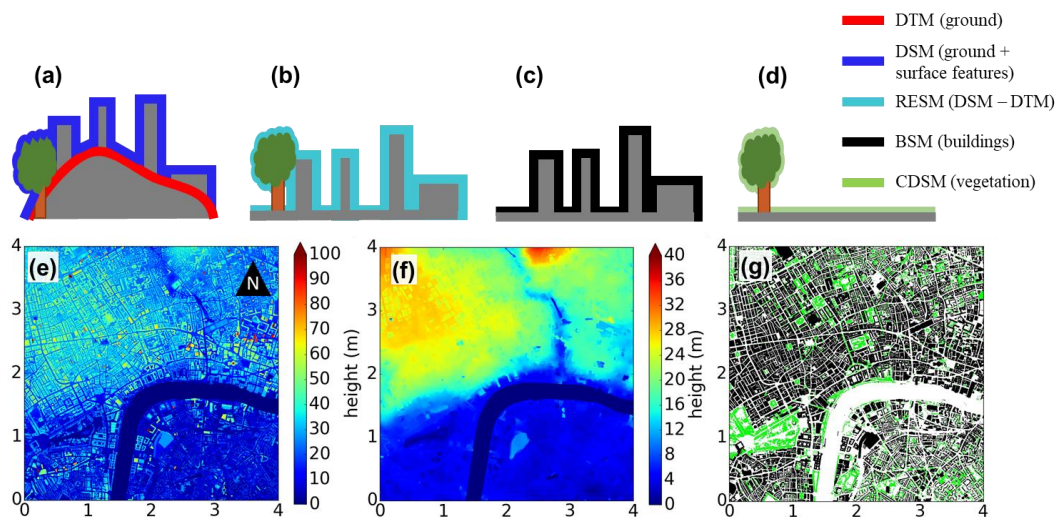
The objective of this work is to assess global digital elevation models (GDEMs) for the determination of urban morphology. First, GDEMs from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Shuttle Radar Topography Mission (SRTM) and TanDEM-X (Table 1) are compared to benchmark elevation data in central London, derived from airborne light detection and ranging (lidar). As quantitative and qualitative analyses find the geometric and aerodynamic parameters to be more accurate with the TanDEM-X model, further comparison is undertaken in five other cities (Table 2). Using these strategically, globally located cities, attempts are made to optimise geometry and aerodynamic roughness parameters derived from the TanDEM-X. To independently assess roughness parameters determined from the different elevation models with morphometric methods, the associated wind-speed profile estimates aloft central London are compared to those observed with Doppler lidar.

## 2 Elevation data and urban morphology

Digital elevation model (DEM) is a collective term for 2-D surface elevation data (Hirt 2015), including both digital surface models (DSMs) (ground and roughness-element heights) and digital terrain models (DTMs) (ground heights only) (Fig. 1).

Subtraction of a DTM from a DSM provides the roughness-element heights (a roughness-element surface model, RESM). Various procedures may be used (e.g. Goodwin et al. 2009, Lindberg and Grimmond 2011a, Crawford et al. 2016, Kent et al. 2017b) to identify and extract building and vegetation heights, producing a building digital surface model (BSM) and canopy digital surface model (CDSM), respectively (Fig. 1).

High resolution DEMs may accurately resolve surface features to horizontal resolutions of 1 m or less (e.g. Fig. 1e). However, expensive collection, storage and processing means these data tend to be available only locally and the varying methods and/or reporting makes up-to-date global elevation data (currently) unfeasible from such sources.



**Figure 1: Digital elevation models (DEMs) & terminology:** (a) digital surface model (DSM) and digital terrain model (DTM), (b) roughness-element surface model (RESM = DSM - DTM), (c) building digital surface model (BSM) & (d) canopy digital surface model (CDSM). From (b) to (c) or (d) requires land cover information. GIS examples are: (e) DSM, (f) DTM, (g) pixels of the BSM (black) and CDSM (green) for a 4 km x 4 km area in central London, UK (map units: km). Surface elevation database details: Lindberg & Grimmond (2011a).

**Table 1: Summary of the global digital elevation models used in this work.**

DEM	Spatial extent	Temporal extent	Method of creation	Horizontal resolution <sup>a</sup>	Horizontal accuracy <sup>b</sup>	Vertical accuracy <sup>b</sup>	Datum: horizontal, vertical	Data source <sup>c</sup>	Cost	Key references
ASTER (Version 2)	83° N & S	2000 – 2010	Photogrammetry	1 arc-second	<6 m <sup>d</sup>	<15 m (95% confidence)	WGS84 EGM96	USGS <sup>e</sup>	Free	ASTER GDEM Validation Team (2009), Tachikawa et al. (2011)
SRTM 1 Arc-Second Global	60° N to 56° S	11 – 22 Feb 2000	Interferometry		<12.6 m (90% confidence) <sup>f</sup>	<9.0 m (90% confid.) <sup>f</sup>				Farr and Kobrick (2000), Farr et al. (2007)
TanDEM-X	Global	Dec 2010 – Jan 2015	Interferometry	0.4 arc-second	<10 m (90% confidence)		WGS84 WGS84	DLR <sup>g</sup>	Free scientific grant or €10/ km <sup>2</sup>	Wessel (2016), Rizzoli et al. (2017)

<sup>a</sup> 1 arc-second is approximately 30 m at the equator <sup>b</sup> As reported during internal evaluation (see key references) <sup>c</sup> As used during this work

<sup>d</sup> 0.13 arc-sec. to west, 0.19 arc-sec. to north (Tachikawa et al. 2011) <sup>e</sup> US Geological Survey data explorer (<https://earthexplorer.usgs.gov/>, accessed Nov. 2016)

<sup>f</sup> For spatial variability see Rodriguez et al. (2006) <sup>g</sup> Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Centre, research grant: DEM\_URBAN1021, data provided: February 2017)

Globally, topographic data can be obtained from orbital SAR (e.g. SRTM and TanDEM-X) or optical stereographic images (e.g. ASTER) (Bürgmann et al. 2000, Stevens et al. 2004). Although, other GDEMs exist (e.g. Rexer and Hirt 2016), the ASTER, SRTM and TanDEM-X models are assessed here given their extensive use and accessibility (Table 1). Typically, GDEM vertical and horizontal accuracy is evaluated against ground control points (e.g. Li et al. 2013) or a higher accuracy elevation model (Hofman et al. 2006, Guth 2010, Tachikawa et al. 2011, Ioannidis et al. 2014). In this work, the latter approach is used.

Although many remote sensing studies explore urban land cover change, few have assessed the ability of GDEMs to reproduce the 3-D structure of cities (Yang et al. 2011). Cities complex structure is problematic for remote sensing because of layover, shadowing and multipath artefacts (Farr et al. 2007, Sportouche et al. 2011, Small and Sohn 2015, Wang et al. 2018). This means individual roughness elements tend to have ‘indistinct’ boundaries (e.g. Eckert and Hollands 2010, Zeng et al. 2014, Xu et al. 2017), especially as resolution coarsens and becomes less than roughness-element dimensions and details (e.g. roof pitch, different roof heights, etc.). However, these inaccuracies may remain even with higher resolution data. For example, comparing  $\leq 3$  m horizontal resolution satellite-derived stereo and SAR imagery to benchmark data in Hong Kong, Xu et al. (2017) demonstrate imperfect building boundaries, with height underestimation in high-density locations and potential overestimation of lower buildings. The urban morphology parameters: building coverage, building height, frontal area index and  $z_0$  (calculated with Macdonald et al. 1998 morphometric method) had absolute average differences of between 22 and 30%, compared to benchmark data.

Other analysis indicates tall buildings locations may be accurate in the SRTM data, but heights are underestimated (e.g. Gamba et al. 2002, Small and Sohn 2015). Similarly, for the TanDEM-X in Berlin, Rossi and Gernhardt (2013) found underestimation of roughness-element heights (buildings and trees) and also overestimation of ground heights between them. Marconcini et al. (2014) and Geiß et al. (2015) provide methodologies to extract DTMs from the TanDEM-X model, but without quantitative assessment of the resulting roughness-element geometry. Automated analysis of the TanDEM-X data has generated a binary global urban footprint (GUF) mask, delineating ‘settlement’ or ‘non-settlement’ areas (Esch et al. 2011, 2013, 2017).

Combining ASTER, Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) and night-time light images, Darmanto et al. (2017) derive empirical relations to obtain plan area (Fig. 3) and average heights of buildings in three cities (Istanbul, Tokyo, Jakarta). These are used with Tokyo-based empirical relations (Kanda et al. 2013) to calculate aerodynamic roughness parameters ( $z_d$  and  $z_0$ ). The roughness parameters are not quantitatively assessed, but provide more accurate wind-speed estimates (cf. observations) than using default settings in a single-layer urban canopy model coupled with the weather research and forecasting model (SLUCM/WRF).

### 3 Methodology

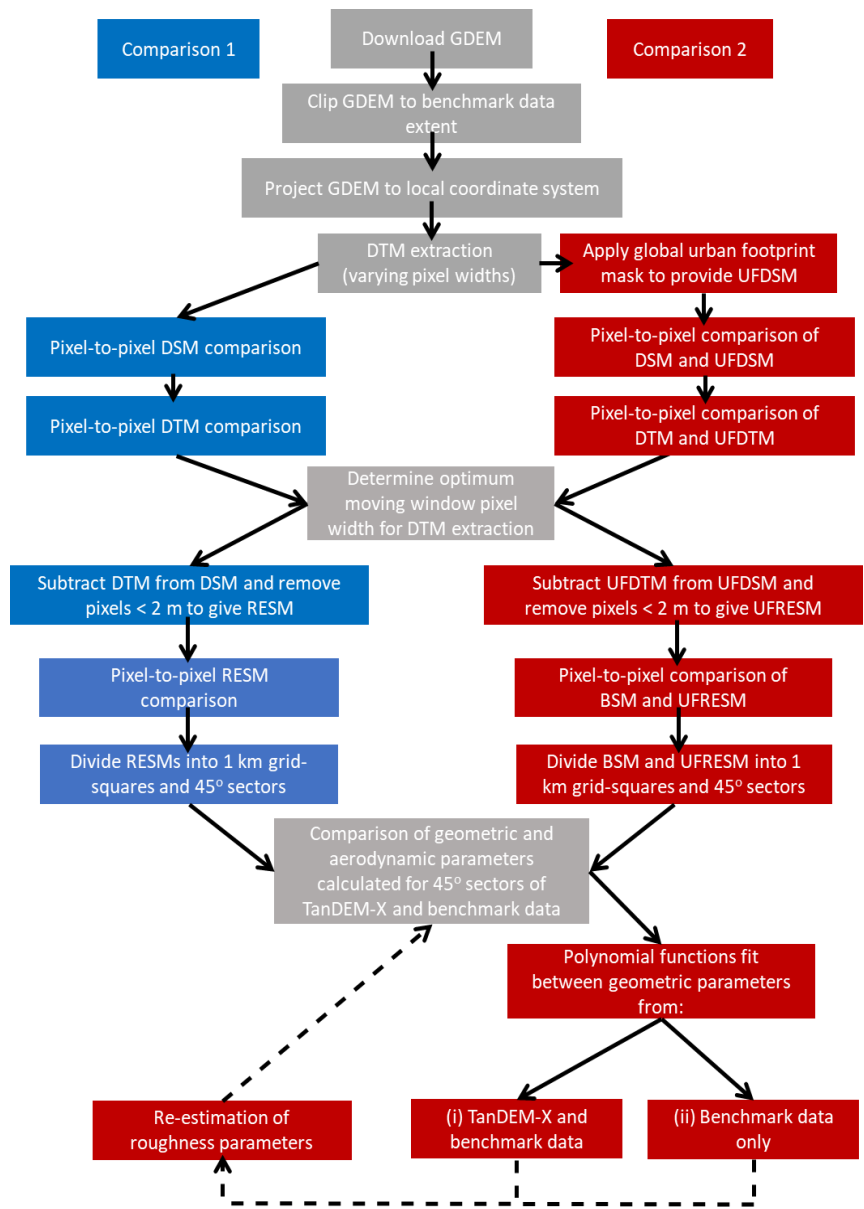
#### 3.1 Dataset comparison

The analysis is conducted in two stages (Fig. 2): Comparison 1 – ASTER, SRTM and TanDEM-X are compared to benchmark data for a 20 km x 20 km area in central London; and, Comparison 2 – TanDEM-X is compared to benchmark data in five cities (Table 2). Note, benchmark data are not the ‘truth’ as these also have uncertainties (e.g. the precise location and height of buildings). All comparisons are performed in the horizontal datum local to the benchmark data (Table 2). The TanDEM-X heights are referenced to the WGS84 ellipsoid and therefore converted to orthometric height using the National Geospatial intelligence agency’s 2.5 x 2.5-minute horizontal resolution offset between the WGS84 ellipsoid and EGM2008 Geoid (NGA 2008). Each of the assessed GDEMs are sensitive to all surface features, including vegetation (e.g. Farr et al. 2007, Tachikawa et al. 2011, Wessel 2016, Rexer and Hirt 2016). The GDEMs are therefore assumed to be DSMs and subtraction of ground heights (DTM) provides a roughness-element surface model (e.g. Fig. 1) from which morphology parameters are calculated. The method to extract ground heights (DTMs) is given in Sect. 3.2.

Initially, a pixel-to-pixel comparison between the GDEMs and benchmark data is undertaken.

Although not central to determining the morphology parameters, the comparison informs the DTM extraction (Sect. 3.2) and provides insight to the GDEM’s ability to reproduce intricacies of the urban surface. For results see Appendix B.

The RESM created from each elevation dataset is used to calculate geometric and roughness parameters. Parameters are calculated using 1 km x 1 km grid-squares with a 500-m overlap, subdivided into eight 45° sectors or wind directions (Fig. 3). Directional sectors allow for the variability of urban morphology with associated upwind roughness to be included (e.g. a site’s fetch may vary with prevailing wind). The 500-m overlap ensures the entire surface area in the DEMs are considered for each direction. The upwind distance and directional sector width are based upon measurement source areas for observations during neutral atmospheric stability at approximately 2.5 times the canopy height in central London (e.g. Fig. 2.3). The scale for calculations is also consistent with the grid size used to derive  $z_d$  and  $z_0$  for the Kanda et al. (2013) morphometric method, as well as spacing employed in commonly used urban modelling systems, e.g. within the Weather Research and Forecasting (WRF) model (Chen et al. 2011).



**Figure 2:** Processing work flow of the global digital elevation models (GDEMs) and for comparison to benchmark data. For abbreviations see Fig. 1. For comparison 2, only pixels within the global urban footprint (GUF) mask ‘settlement layer’ are retained (Esch et al. 2017), indicated with ‘UF’ prefix: UFDSM – urban footprint digital surface model = ground + surface feature heights within the GUF; UFDTM – urban footprint digital terrain model = ground heights within the GUF; UFRESM – urban footprint roughness-element surface model = roughness-element heights within the GUF.

The RESM is used to calculate geometric and roughness parameters for each elevation dataset. Parameters are calculated using 1 km x 1 km grid-squares with a 500-m overlap, subdivided into eight 45° sectors or wind directions (Fig. 3). Directional sectors allow for the variability of urban morphology with associated upwind roughness to be included (e.g. a site’s fetch may vary with prevailing wind). The 500-m overlap ensures the entire surface area in the DEMs are considered for each direction. The upwind



distance and directional sector width are based upon measurement source areas for observations during neutral atmospheric stability at approximately 2.5 times the canopy height in central London (e.g. Fig. 2.3). The scale for calculations is also consistent with the grid size used to derive  $z_d$  and  $z_0$  for the Kanda et al. (2013) morphometric method, as well as spacing employed in commonly used urban modelling systems, e.g. within the Weather Research and Forecasting (WRF) model (Chen et al. 2011).

For each sector, the geometric parameters determined are the average ( $H_{av}$ ), maximum ( $H_{max}$ ) and standard deviation ( $\sigma_H$ ) of roughness-element heights, the plan area index ( $\lambda_p$ ) and the frontal area index ( $\lambda_f$ ) (Fig. 3c). Only roughness elements with height  $> 2$  m are considered, ensuring low-level street furniture (e.g. signage), vehicles, etc. are removed. The parameters are used with Kanda et al.'s (2013) morphometric method to calculate  $z_d$  and  $z_0$ , which directly incorporates roughness-elements height variability and has been found to provide more accurate wind-speed estimates than other methods (Varquez et al. 2015, Kent et al. 2017a, 2017b, 2018). The average ground height ( $H_{av,grd}$ ) in each sector is recorded from the DTM.

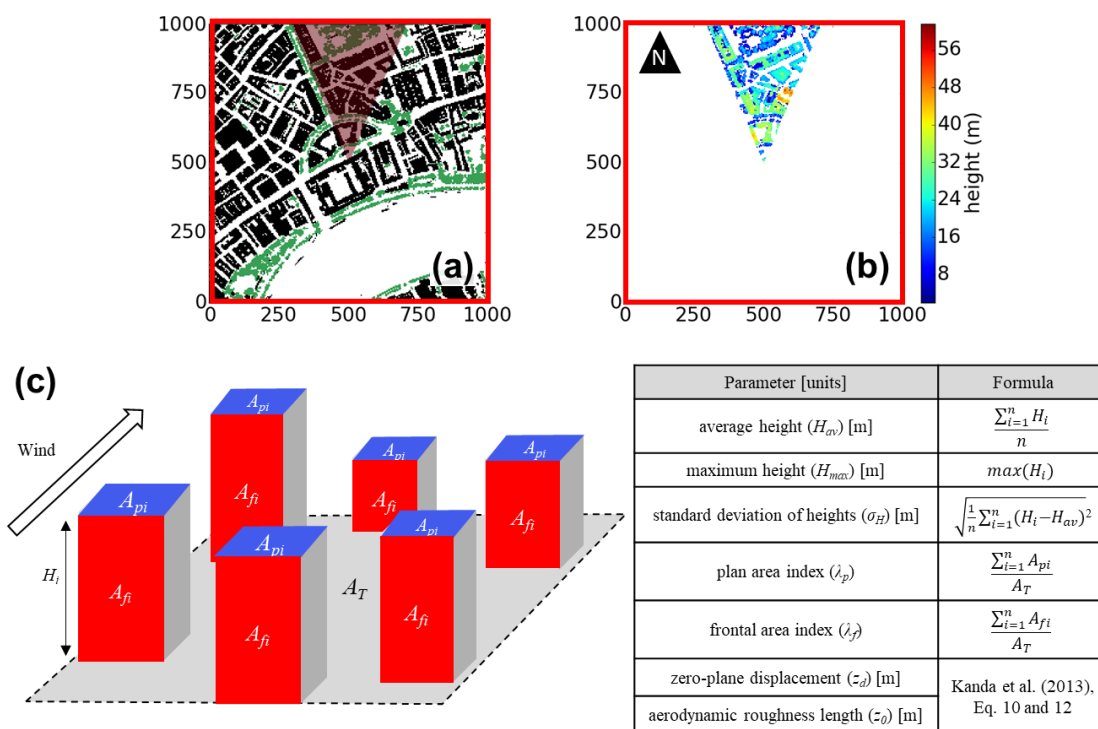
In comparison 1, the benchmark data allows all roughness-element heights to be considered. However, only built roughness-elements are used in comparison 2 as some benchmark datasets only contain building heights. As with previous studies (e.g. Marconcini et al. 2014), the GUF mask is applied across the entire comparison domain to retain building heights from the TanDEM-X model (indicated with 'UF', i.e. UFDMS, UFDTM and UFDRESM). In Auckland, the absence of benchmark building footprints data required the GUF mask to be applied to both the benchmark and TanDEM-X data. Insight to the roughness elements captured by the GUF mask is provided in Appendix A. To permit comparison to previous analyses, the metric used to assess the results is the root-mean-square error (RMSE) with the normalised RMSE (nRMSE = RMSE divided by mean of 'observed' values) allowing for a scale-independent comparison of errors.

**Table 2:** Source and extent of the benchmark data used in comparison 2 (Fig. 2). Benchmark datasets downloaded during March 2017.

Symbols: ✓ = available from source, - = not available, C = created. Maps of spatial extent provided in Fig. 10 (New York) and supplementary material.

City	EPSG (WGS84 UTM zone)	Upper left co-ordinates ( $x_{min}$ , $y_{max}$ )	x extent, y extent (km)	DSM	DTM	BSM <sup>a</sup>	CDSM	Resolution	Collection method	Collection date	Source
Sao Paulo	32723 (23S)	313901.002, 7411738.415	47, 57	-	-	✓ (abs)	-	Building footprint	Photogrammetry	2007, updated yearly	GeoSampa (2017), Danilo Mizuta pers. comm. (16/8/2017)
Tokyo	32654 (54N)	364005.246, 3966451.299	46, 38	-	✓	✓ (st)	-	DTM: 5 m, BSM: building footprint	DTM: lidar BSM: ground survey	2011	Tokyo Metropolitan GIS Data for Urban Planning
New York	32618 (18N)	563127.939, 4529850.727	47, 46	-	✓	✓ (abs)	-	DTM: 1 m, BSM: building footprint	DTM: Leica ALS70 lidar BSM: aerial imagery	DTM: Mar-Apr 2014 BSM: 2014	DTM: USGS CMGP (2014) BSM: NYC DoITT (2014)
London	32631 (31N)	253226.215, 5732341.367	61, 50	✓	✓	✓	✓	1 m	Optech ALTM 3033 lidar	Aug-Sept 2008	Lindberg and Grimmond (2011a)
Auckland	32760 (60S)	293156.620, 5921972.196	19, 13	✓	✓	C <sup>b</sup>	-	1 m	Optech ALTM 3100 lidar	July-Nov 2013	LINZ (2013)

<sup>a</sup> Method of reporting heights above ground level: (abs) is absolute heights, (st) is storey based heights. Correction from (st) to (abs) uses the Council on Tall Buildings and Urban Habitat (CTBUH 2017) recommendations, with a storey  $s = 3.5$  m, except for 'tall buildings' ( $> 14$  stories,  $H_{i,tall}$ ):  $H_{i,tall} = 3.5s + 9.625 + 2.625(s/25)$  <sup>b</sup> Created from the TanDEM-X global urban footprint mask 'settlement' layer (Esch et al. 2017) (see text for details)



**Figure 3:** Procedure to determine morphology parameters from a directional sector of a 1 km x 1 km grid square in central London (red square in Fig. 4a): (a) roughness-element surface model (black = buildings, green = vegetation) overlain with 45° sector representing northerly wind direction (red shading); (b) roughness-element heights within directional sector for calculation; (c) parameter calculation from idealised roughness elements.

### 3.2 Extraction of ground heights (DTM) from the GDEMs

Sophisticated techniques exist to separate ground heights and surface features from high-resolution point-cloud data (e.g. Kraus and Pfeifer 2001, Chen et al. 2009, Schreyer et al. 2014). However, the comparatively coarse resolution of GDEMs require other techniques. Marconcini et al. (2014) extract DTMs from the TanDEM-X model assuming roughness elements are pixels above a threshold elevation. Threshold-differencing techniques may be inaccurate in regions where the amplitude of ground elevation range exceeds building heights, leading to higher bare-earth pixels being misidentified as buildings. Therefore, Geiß et al. (2015) propose object-orientated, progressive morphological filtering techniques. Neither Marconcini et al. (2014) or Geiß et al. (2015) quantitatively analyse RESMs produced using the extracted DTMs

Here, DTM extraction uses a moving window. The ground height of the central pixel is assumed to correspond to the lowest elevation pixel within the window. Arbitrary moving windows have been used (e.g. 200 m x 200 m, Gamba et al. 2002). Here the pixel-width of the moving square window is varied (3-, 5-, 7-, 11-, 21- and 31-pixels; corresponding to a 1-, 2-, 3-, 5-, 10- and 15-pixel buffer from the central pixel). The six extracted DTMs are compared to the benchmark DTM to identify the most appropriate moving window size. The DTM extraction procedure is swift, objective and responds to local elevation, eliminating requirements to set subjective thresholds during threshold-differencing (e.g. Marconcini et al. 2014), and individual pixel classification during object-orientated approaches (e.g. Geiß et al. 2015).

### 3.3 Polynomial function fitting

During comparison 2, polynomial functions are fit through city-specific data on two occasions (Fig. 2) with the objective of investigating the relation between different geometric parameters determined with the: (i) TanDEM-X and benchmark data and (ii) benchmark data only. Initially locally weighted scatter-plot smoother (LOWESS, Cleveland 1979) data pairs are obtained, which are used in a non-linear least squares fit (Bates and Watts 1988) constrained to the polynomial function:

$$y = ax^3 + bx^2 + cx \quad (1)$$

where  $a$ ,  $b$  and  $c$  are best-fit coefficients for the geometric parameters  $x$  and  $y$  listed in Table 3. Equation 1 permits a variable relation across the range of data, while ensuring if  $y = 0$  then  $x = 0$ . Using the LOWESS data intentionally removes the sensitivity to outliers (e.g. if area has changed over time, Sect. 6). To ensure the fits have sufficient data, the  $x$  parameter data are binned (5 m for  $H_{av}$  and  $\sigma_H$ , 10 m for  $H_{max}$ , and 0.1 for  $\lambda_p$  and  $\lambda_f$ ) and fits restricted to bins  $n \geq 20$ . The Akaike (1974) information criterion indicates the selected order of polynomial (Eq. 1) does not lead to overfitting of the data. Generating ‘multi-city’ polynomial functions with data points from all cities leads to bias towards cities with more data. Therefore, multi-city functions are created with the same fitting procedure as above, but through 1000 evenly sampled points along city-specific polynomials. With this number of points a  $\pm 10\%$  change (i.e. using 900 or 1100 points) does not impact the fitted coefficients in Eq. 1.

(a)		(b)	
$x$ [TDX]	$y$ [BM]	$x$ [BM]	$y$ [BM]
$H_{av,grd}$	$H_{av,grd}$	$\lambda_{p,b}$	$\lambda_{f,b}$
$H_{av}$	$H_{av}$	$H_{av,b}$	$\sigma_{H,b}$
$H_{max}$	$H_{max}$	$\sigma_{H,b}$	$H_{max,b}$
$\sigma_H$	$\sigma_H$		
$\lambda_p$	$\lambda_p$		
$\lambda_f$	$\lambda_f$		

**Table 3:** Geometric parameters  $x$  and  $y$  used during polynomial function fitting (Eq. 1). Polynomial fits are between geometric parameters derived from the: (a) TanDEM-X [TDX] and benchmark [BM] elevation data (i.e. fitting stage (i) in Fig. 2) and (b) benchmark parameters only (i.e. fitting stage (ii) in Fig. 2). Parameter calculation, abbreviations and units: Fig. 3c.

## 4 Results

### 4.1 Comparison 1: ASTER, SRTM and TanDEM-X in London, UK

Initial comparison of the ASTER, SRTM and TanDEM-X GDEMs for a 20 km x 20 km area in central London provides insight to their respective representations of the urban surface compared to high-resolution (benchmark) data (Fig. 4). The similarity of SRTM and TanDEM-X elevations (and therefore differences from the benchmark data) are likely due to their derivation from similar interferometric methods. Elevations from the photogrammetric approach used to derive ASTER appear noisy, producing the greatest differences from the benchmark data (cf. Fig. 4c – f). The differences for the ASTER data can be considerable ( $\pm 20$  m) across the study area, whilst differences for the SRTM and TanDEM-X data become largest in the city centre due to the more complex geometry.

For the pixel-to-pixel comparison (Appendix B) the RMSE assessed against the benchmark data for the DSMs have values of 9.79 m (ASTER), 6.27 m (SRTM) and 6.35 m (TanDEM-X), which are all better than the accuracy given by respective agency studies (Table 1). The errors are largest for the greatest surface heights, due to the underestimation of taller roughness elements. Topographical variability also contributes to larger errors in ASTER (Fig. 4f). Underestimation of roughness-element heights cannot be attributed to the DTM extraction procedure (see Appendix B for discussion), for which a 5 x 5-pixel moving window extracts DTMs which best resemble the benchmark data for all GDEMs (RMSE  $\sim 3$  m for TanDEM-X and SRTM, and 8 m for ASTER). Despite the ‘block-like’ DTM which is extracted from the GDEMs (Fig. 6f-h), multiple iterations of the procedure or smoothing (e.g. Gaussian filters and interpolation techniques) were not found to improve the pixel-based accuracy of the method.

In total, 12800 samples for each geometric and aerodynamic parameter are compared (20 km x 20 km study area = 1600 overlapping grid-squares each with 8 directions). Consistent underestimation of roughness-element heights by the GDEMs causes the height based geometric parameters ( $H_{av}$ ,  $H_{max}$  and  $\sigma_H$ ) to be increasingly underestimated as roughness-element heights become taller (Fig. 5a-c). The underestimation is likely due to the coarse resolution of the GDEMs causing apparent merging of roughness elements (Fig. 6), which also leads to tendency to overestimate  $\lambda_p$  and underestimate  $\lambda_f$  (Fig. 5e and f). As  $z_d$  requires  $H_{av}$ ,  $\sigma_H$ ,  $H_{max}$  and  $\lambda_p$  (Kanda et al. 2013, their Eq. 10), the  $z_d$  results resemble the height-based parameters (Fig. 5d). However, the overestimation

of  $\lambda_p$  can cause overestimation of  $z_d$ , especially for  $z_d < 10$  m. Combined these effects result in an unreasonably small  $z_0$  (typically  $\sim 10^{-3}$  m) with the largest relative error (nRMSE) from the benchmark data (Table 4). The TanDEM-X data best resembles the benchmark data for all parameters (Table 4).

Results from one grid-square (Fig. 6) showcases the apparent merging and underestimation of roughness-element heights by each GDEM (Fig. 6a-d). Merging can lead to overestimation of the ground height between roughness elements (Fig. 6e-h) and a lack of spaces between individual roughness elements (e.g. buildings). These effects are less pronounced for the TanDEM-X data, hence the better performance relative to benchmark data. The width and location of the river (south of grid-square) is accurately resolved by the TanDEM-X water mask (Appendix A). The SRTM resolves the river but with a smaller width (cf. Fig. 6i and l), whilst the river's presence is not obvious in the ASTER (Fig. 6k).

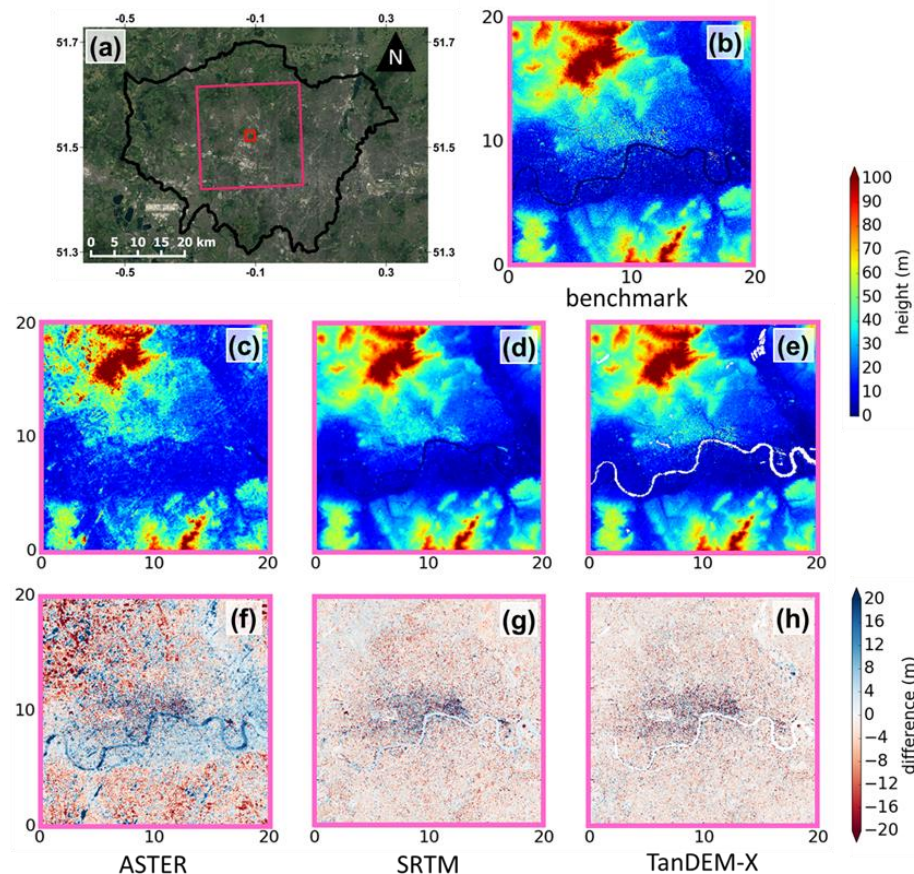
#### 4.2 Comparison 2: TanDEM-X in five cities

##### 4.2.1 Comparison to benchmark data

Pixel-to-pixel comparison of TanDEM-X and benchmark data in five cities substantiates the comparison 1 findings. However, as more complex geometry is encountered errors become larger, especially in areas with densely packed, taller buildings with a smaller plan area. Like comparison 1, errors result from overestimation of ground heights between buildings (radar signal not penetrating to ground level) and underestimation of taller building heights (up to a factor of 10). Figure 7 demonstrates these effects for Manhattan, New York, with the TanDEM-X clearly a merged representation of the benchmark data (Fig. 7). These effects (and errors) are less in areas of less complex morphology (e.g. Auckland, London). More detailed results are provided in Appendix B.

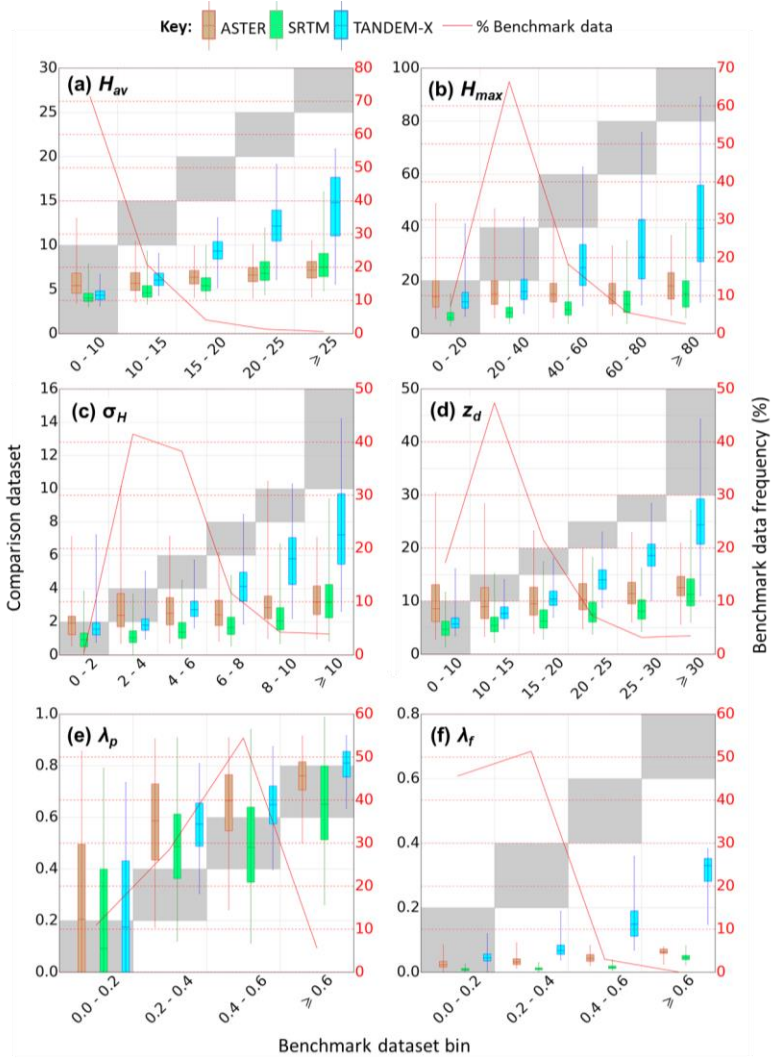
The average ground height is the best estimated parameter ( $RMSE < 2$  m;  $nRMSE < 0.1$ ) (Fig. 8a, Table 5). Consistent underestimation of height-based parameters gives larger errors that increase with height (Fig. 8b-d). The apparent merging of roughness elements causes overestimation of  $\lambda_p$  (Fig. 8e) and underestimation of  $\lambda_f$  (Fig. 8f).

Polynomial fits (Sect. 3.3) between the benchmark and TanDEM-X parameters (Fig. 8, coloured lines) vary for each city because of inter-city morphological variability and the associated variable performance of the TanDEM-X data. For example, the more high-rise nature of buildings in New York creates increased potential for layover and shadowing effects and larger estimates of  $\lambda_p$ , which require greater correction compared to the other cities (Fig. 8e). The scatter of data around the polynomial fits for each city (i.e. inter-city variability) is associated with similar effects, but between different neighbourhoods of a city (e.g. downtown compared to suburban). The multi-city fits provide a starting point for the expected relation across cities globally, but detailed investigation of GDEM performance in different neighbourhoods of the city will contribute to quantifying uncertainty.



**Figure 4:** (a) Greater London (black demarcation), with the 20 km x 20 km study area (magenta) and a central area (red) with meteorological equipment (Sect. 5). The digital surface model (ground + roughness-element heights, DSM) of the (b) benchmark is shown, with DSMs from the (c) ASTER, (d) SRTM, and (e) TanDEM-X (with water mask applied [white]) and (f–h) their respective differences from the benchmark DSM. Map units: (a) degrees (WGS84) and (b–h) km.

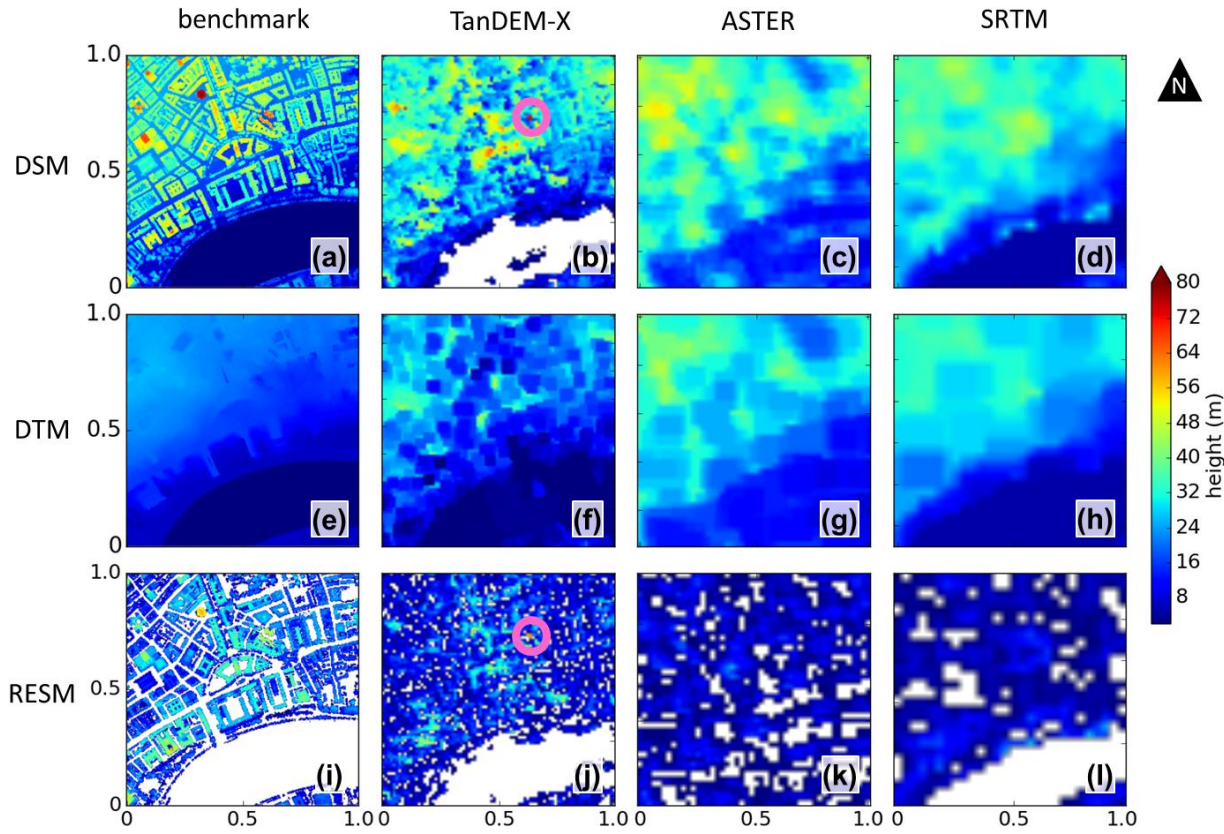




**Figure 5:** Comparison between parameters determined from the GDEMs and benchmark data for eight directional sectors of 1 km grid-squares in the 20 km x 20 km central London area (Fig. 4). For the comparisons, parameters in the benchmark dataset are divided into in to bins (x-axis, not inclusive of upper value labelled) and the distributions of parameters calculated by the GDEMs in the corresponding locations are shown for each bin (left y-axis). Whiskers are the 1<sup>st</sup> and 99<sup>th</sup> percentiles. Note, ideal distributions would be within the grey shaded areas. The percentage frequency of benchmark values within each bin are also shown (right y-axis, red line). Parameter calculation, abbreviations and units: Fig. 3c.

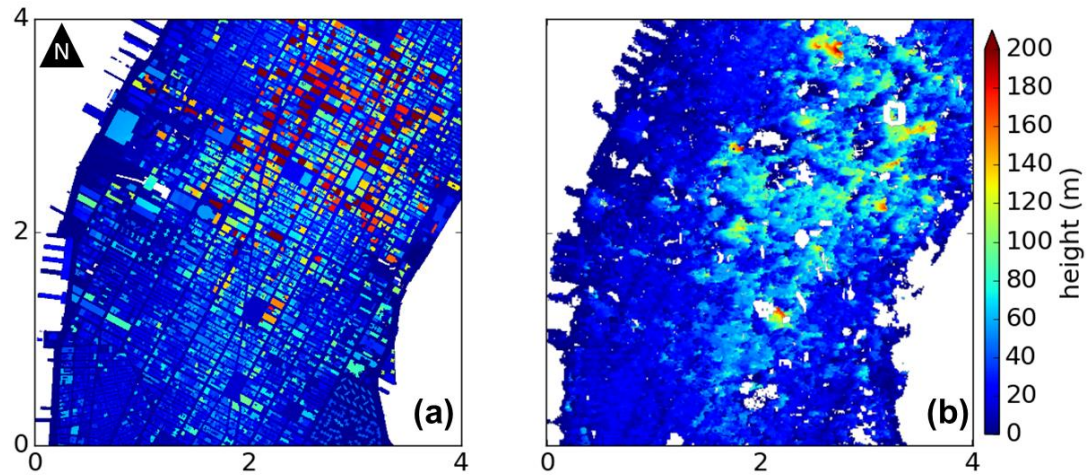
**Table 4:** Root-mean-square error (RMSE) and normalised RMSE (nRMSE) for geometric and aerodynamic parameters calculated from the benchmark data and the: ASTER, SRTM and TanDEM-X datasets. Parameters are calculated for eight directional sectors of 1 km grid-squares in the 20 km x 20 km central London area (Fig. 4). For each parameter, errors are calculated if both elevation models have  $H_{av}$  and  $H_{max} > 2$  m and the benchmark data have  $\lambda_p$  and  $\lambda_f \geq 0.05$  and  $z_d$  and  $z_0 > 0.1$ . Parameter calculation, abbreviations and units: Fig.3c.  $H_{av,grd}$  is the average ground height [m]. Note similarity in  $z_0$  errors for ASTER and SRTM is due to rounding.

Parameter	ASTER		SRTM		TanDEM-X	
	RMSE	nRMSE	RMSE	nRMSE	RMSE	nRMSE
$H_{av,grd}$	5.87	0.23	2.22	0.09	1.43	0.06
$H_{av}$	5.16	0.55	5.75	0.61	4.71	0.50
$H_{max}$	27.64	0.77	31.59	0.88	21.11	0.59
$\sigma_H$	3.95	0.80	4.28	0.87	2.93	0.59
$\lambda_p$	0.26	0.65	0.22	0.55	0.23	0.57
$\lambda_f$	0.21	1.01	0.23	1.10	0.17	0.82
$z_d$	9.33	0.63	10.37	0.70	6.65	0.45
$z_0$	0.92	1.58	0.92	1.58	0.89	1.53



**Figure 6:** 1 km grid-square in central London (red square in Fig. 4a): (a-d) digital surface model (DSM), (e-h) digital terrain model (DTM) (for the GDEMs extracted using the Sect. 3.2 methodology with a 5-pixel moving window) and (i-l) roughness-element surface model (RESM) (rows) derived from the: benchmark, TanDEM-X, ASTER and SRTM datasets (columns). Magenta circles are referred to in text (Sect. 5). Map units: km.

**Figure 7:** Digital surface models of downtown New York from the (a) benchmark and (b) TanDEM-X data. Only buildings and ground heights are included in the benchmark data. The global urban footprint (GUF) mask (Esch et al. 2017, see text) is applied to the TanDEM-X model. Map units: km.



#### 4.2.2 Empirical correction of parameters

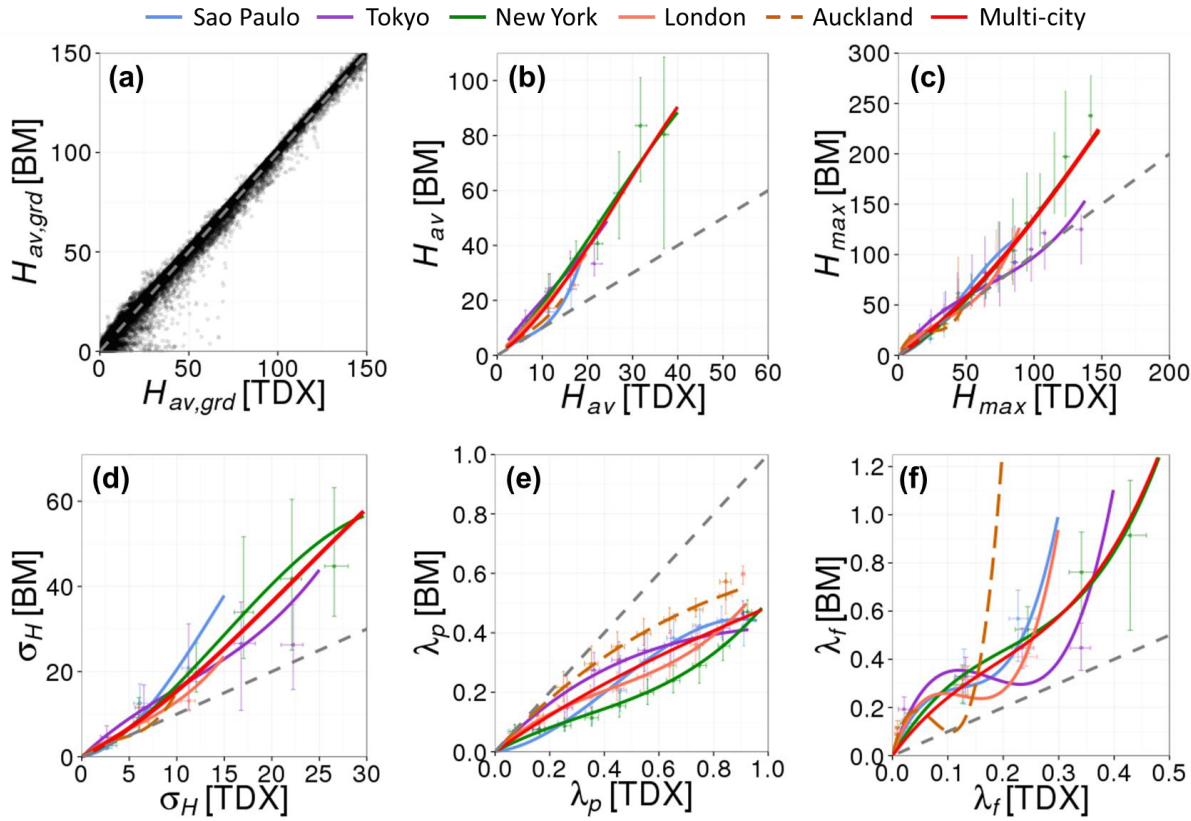
City-specific and multi-city polynomials are used to correct the geometric parameters derived from the TanDEM-X data. Both correction methods provide improved parameter estimates. City-specific corrections are slightly better, with the RMSE reduced for height-based parameters ( $H_{av}$  (35%),  $H_{max}$  (10%), and  $\sigma_H$  (15%)) and for  $\lambda_p$  (70%) (Table 5). However, the spread of  $\lambda_f$  data points results in city-specific fits being highly variable (Fig. 8f) and the multi-city corrections (Fig. 8, red line) produce less error (Table 5). Even after the  $\lambda_f$  is corrected with the multi-city fit, considerable error may exist in smaller  $\lambda_f$  values due to the large range of differences. Testing reveals estimation of these smaller  $\lambda_f$  values is improved through using the corrected  $\lambda_p$  and the multi-city relation found between  $\lambda_f$  and  $\lambda_p$  (Appendix D: Fig. D1a, red line):

$$\lambda_f = 0.46\lambda_p^3 - 0.39\lambda_p^2 + 0.55\lambda_p \quad (2)$$

The RMSE is minimised if the additional correction (Eq. 2) is applied to corrected values of  $\lambda_f < 0.08$ .

Aerodynamic roughness parameters determined across the five cities with uncorrected geometry from the TanDEM-X data are consistent with comparison 1 results. Underestimation of  $z_d$  increases as benchmark  $z_d$  increases (Fig. 9a) and  $z_0$  values are unreasonably small (Fig. 9b). If the TanDEM-X geometric parameters are corrected, and then used to calculate roughness parameters, there are reductions in RMSE for both  $z_d$  (20%) and  $z_0$  (25%) (Fig. 9c-f, Table 5).





**Figure 8:** Geometric parameters determined for five cities based on 1 km grid-squares (8 directional sectors per grid, Fig. 3) from benchmark [BM] and TanDEM-X [TDX] datasets. All data points are shown for (a) average ground height ( $H_{av,grd}$ ). For other parameters (b-f), each city's data are binned (5 m for  $H_{av}$  and  $\sigma_H$ ; 10 m for  $H_{max}$ ; 0.1 for  $\lambda_p$  and  $\lambda_f$ ) with the median (point) and interquartile range (whiskers) per bin shown. Polynomial fits use all city-specific data for each parameter (colored lines). See Sect. 3.3 for 'multi-city' fit method and text for explanation of fits in (f). For equation, error and data range of each fit see Appendix C. Parameter calculation, abbreviations and units: Fig. 3c.

**Table 5:** Root-mean-square error (RMSE) and normalised RMSE (nRMSE) of TanDEM-X parameters from the benchmark datasets in all cities using the: uncorrected TanDEM-X data, multi-city polynomial correction, and city-specific polynomial correction. No correction is attempted to the ground height ( $H_{av,grd}$ ). For each parameter, errors are calculated if both elevation models have  $H_{av}$  and  $H_{max} > 2$  m and the benchmark data have  $\lambda_p$  and  $\lambda_f \geq 0.05$  and  $z_d$  and  $z_0 > 0.1$ . Parameter calculation, abbreviations and units: Fig. 3c.

Parameter	Uncorrected data		Multi-city polynomial correction		City-specific polynomial correction	
	RMSE	nRMSE	RMSE	nRMSE	RMSE	nRMSE
$H_{av,grd}$	1.95	0.06	-	-	-	-
$H_{av}$	6.29	0.74	4.76	0.56	4.25	0.50
$H_{max}$	18.10	0.72	17.57	0.70	16.77	0.67
$\sigma_H$	4.58	1.03	4.04	0.91	3.91	0.88
$\lambda_p$	0.28	1.04	0.10	0.37	0.08	0.30
$\lambda_f$	0.14	0.97	0.11 <sup>a</sup>	0.76 <sup>a</sup>	0.22	1.52
$z_d$	7.57	0.72	6.71	0.64	6.27	0.60
$z_0$	1.80	2.29	1.34	1.71	1.35 <sup>b</sup>	1.78 <sup>b</sup>

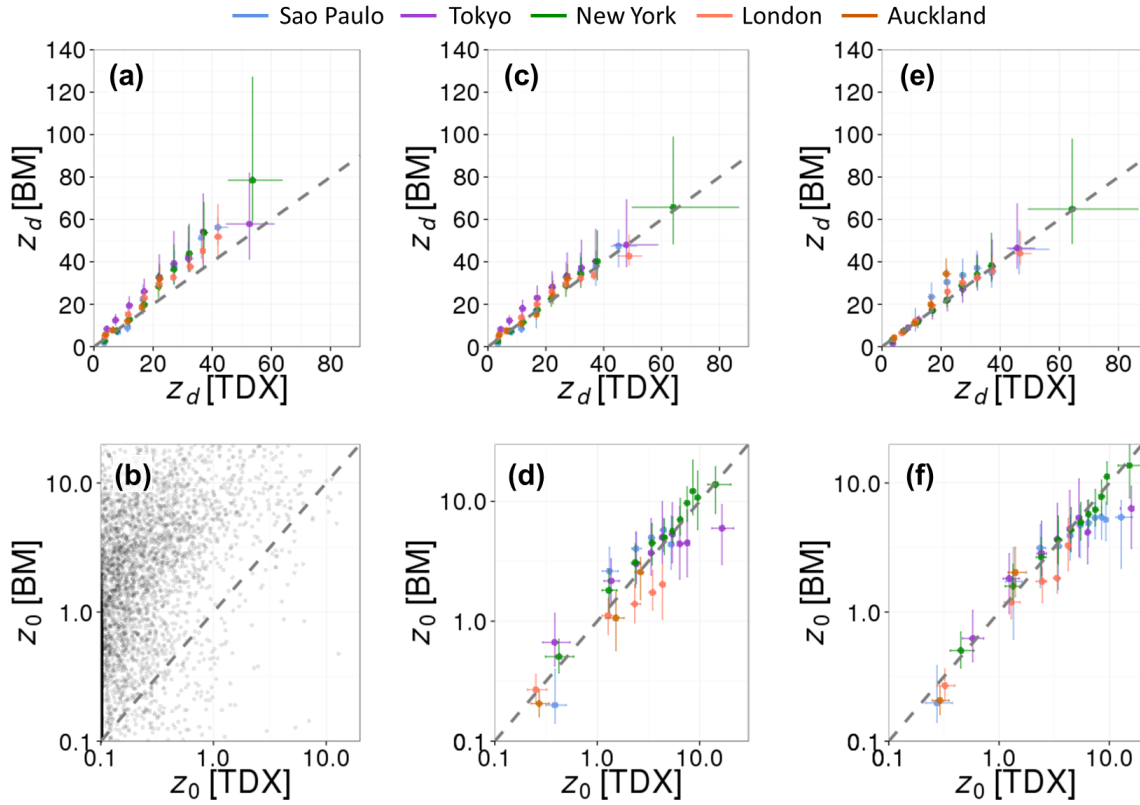
<sup>a</sup> Multi-city correction of  $\lambda_f$  uses Eq. 2 for corrected values of  $\lambda_f < 0.08$  (see discussion in text)

<sup>b</sup> Calculation of the city-specific  $z_0$  uses  $\lambda_f$  corrected with the multi-city function (Fig. 8f, red line)

Numerous other corrections for parameters were explored (e.g. combining relations, data binned by other parameters such as  $H_{av}$  and use of TanDEM-X meta-data layers [not shown]), but hampered by compounding errors and without greater skill relative to the suggested method, i.e.:

- calculate (uncorrected) morphological parameters from the UFRESM
- correct these using city-specific polynomial relations for the most similar city (e.g. for Beijing, Tokyo relations would be used), except for  $\lambda_f$ , which is corrected with the multi-city polynomial relation, but with Eq. 2 for  $\lambda_f < 0.08$ . Alternatively, the multi-city relations can be used with the city-specific corrections offering some range of uncertainty.
- use corrected geometry to calculate  $z_d$  and  $z_0$ .

Using the proposed corrections in the test cities leads to improved estimates of parameters on a city-wide scale (Fig. 10, suppl. material). Additionally, benchmark land cover data from London and New York indicates the 'settlement' and 'non-settlement' TanDEM-X GUF mask contains approximately 70% impervious and pervious surfaces, respectively, whilst the water mask captures up to 75% of water bodies (Appendix A). In combination (e.g. Fig. 10 and Fig. 11), a wide number of parameters required within urban land surface models (e.g. Grimmond et al. 2010, 2011, Salamanca et al. 2011, Varquez et al. 2015) are available. The ability of the corrections to improve wind-speed estimates is demonstrated in Sect. 5.



**Figure 9:** Zero-plane displacement ( $z_d$ ) [m] and aerodynamic roughness length ( $z_0$ ) [m] (note log axis), calculated using the Kanda et al. (2013) morphometric method for five cities based on 1 km grid-squares (8 directional sectors per grid, Fig. 3) from benchmark [BM] and TanDEM-X [TDX] datasets. All data points are shown in (b) and elsewhere data are binned per city (10 m for  $z_d$ ; 1 m for  $z_0$ ) with the median (point) and interquartile range (whiskers) per bin shown. Parameters are calculated with TanDEM-X geometry which is: (a, b) uncorrected; corrected with (c, d) multi-city relations (Fig. 8, Table C1) or (e, f) city-specific relations (Fig. 8, Table C1). Note, multi-city relation is used to correct  $\lambda_f$  in all cases, with Eq. 2 to estimate corrected values of  $\lambda_f < 0.08$  (see discussion in text). Data point errors given in Table 5.

#### 4.2.3 Assessment of empirical corrections for a separate location

An independent assessment of the TanDEM-X derived parameters is undertaken for Slough (Fig. 12), an urban area west of London (Fig. 4a) where both data sets are available. Although Slough is not as complex as other areas considered in this work (cf. Fig. 7a and Fig. 12b), it has a city centre, an industrial area with warehouses, suburbs, water bodies and areas of vegetation. A 5 x 5-pixel moving window is used for DTM extraction from the TanDEM-X. Subsequently, uncorrected (Fig. 13, black points) and corrected (Fig. 13, blue points) parameters (Fig. 12c) are compared to those derived from a benchmark building footprint and height dataset (Fig. 12b).

Using the recommended correction procedure (London-based empirical relations and multi-city relations for  $\lambda_f$ ) improves the characteristic underestimation of  $H_{av}$  (Fig. 13a) and  $\lambda_f$  (Fig. 13e), and the overestimation of  $\lambda_p$  (Fig. 13d). As  $H_{max}$  and  $\sigma_H$  in this area are in the range where TanDEM-X performs reasonably well (Fig. 8c, d), these corrections only have a small effect (~10% reduction in RMSE). Note, the outlying  $\sigma_H$  points in Fig. 13c are where two chimneys up to 100 m tall are located, which are not evident in the TanDEM-X data. In combination, the corrections improve the estimation of  $z_d$  by ~10% (RMSE from 1.93 m to 1.72 m) and  $z_0$  by ~40% (RMSE from 0.39 m to 0.22 m).

The comparison demonstrates an inherent limitation of the correction procedure. If the uncorrected TanDEM-X parameters are accurate (or unlike what the empirical fits in Fig. 8 suggest) the corrections may enhance error, e.g. when the corrections lead to height-based parameters (Fig. 13a-c) or  $\lambda_f$  (Fig. 13e) being overestimated. However, as these are rare (tail) events there is an overall benefit from applying the correction. Furthermore, some evaluation differences may arise from temporal offset of datasets (Sect. 6).

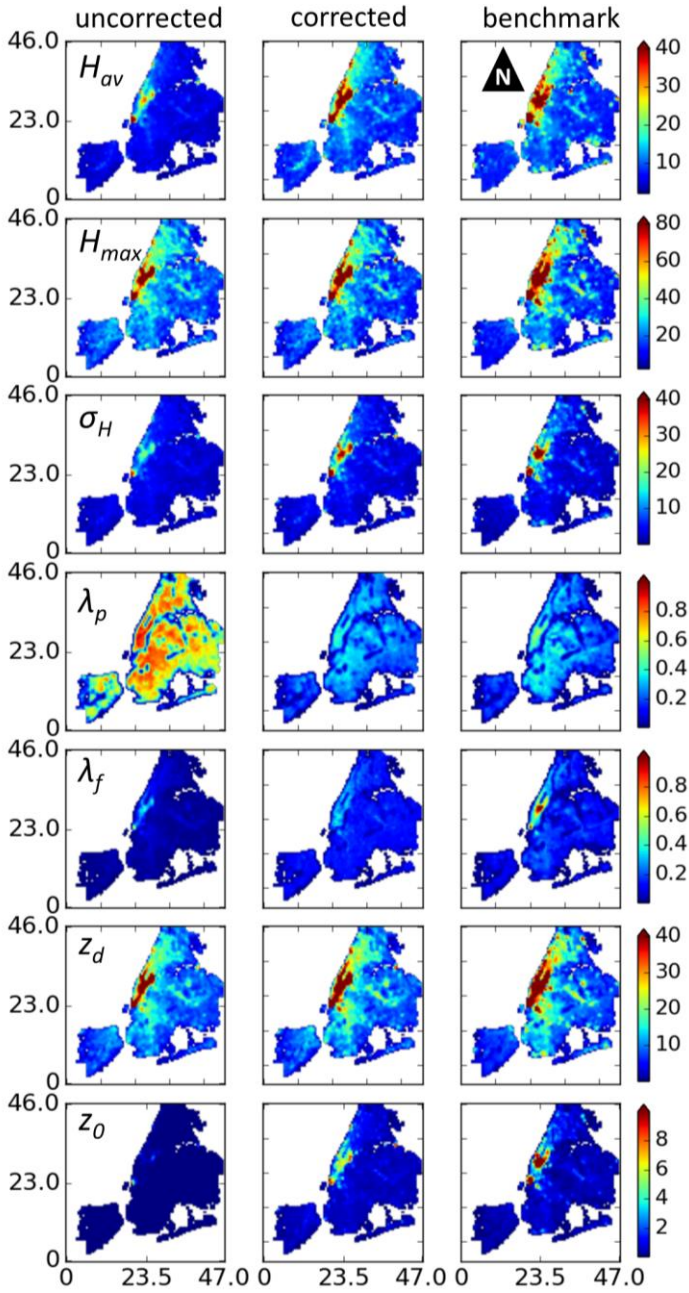
#### 5 Wind-speed estimates using the DEMs

With an average wind speed ( $\bar{U}_{ref}$ ) at a reference height ( $z_{ref}$ ) and roughness parameters ( $z_d$  and  $z_0$ ), the vertical wind-speed profile above a surface can be estimated. To independently assess using roughness parameters derived from the different DEMs, the associated wind-speed estimates are compared to observations in central London. In the centre of the 1 km grid-square in Fig. 6, a sonic anemometer (CSAT3, Campbell Scientific, USA) measured wind speed ( $\bar{U}_{ref}$ ) at approximately 2.5 times the canopy height (49 m above ground level,  $z_{ref}$ ), with a Doppler lidar (Halo Photonics Streamline pulsed) located ~60 m to the west. The latter, operating in doppler beam swinging (DBS) mode, measured the vertical wind-speed profile in 30 m gates aloft. For site, instrument and processing details see Lane et al. (2013), Kotthaus and Grimmond (2014a, b) and Kent et al. (2017a, 2018).

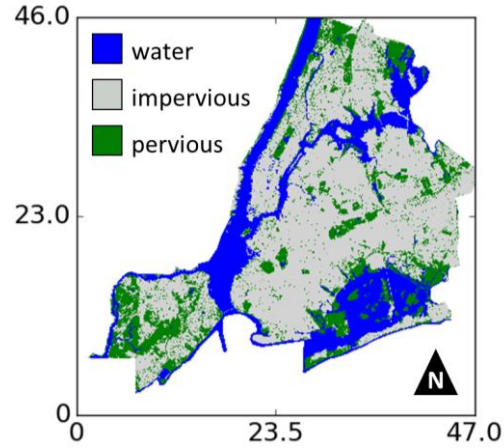
During the observation campaign (Oct 2010 to May 2011), 245 hours of near neutral atmospheric stability occur when it is appropriate to extrapolate the 49-m wind-speed to ~200 m above the canopy without stability corrections (Kent et al. 2018). Here, the same 245 hours are analysed. The roughness parameters calculated from the benchmark and GDEM data are used with the 49-



m wind-speed ( $\bar{U}_{ref}$ ) to extrapolate to the Doppler lidar gate level wind speed. The profile used for extrapolation varies according to wind direction. Following Kent et al. (2018), the logarithmic wind-speed profile (Blackadar and Tennekes 1968) is used for 000° – 045° wind directions, whilst the Deaves and Harris equilibrium profile (Deaves and Harris 1978) is used elsewhere.



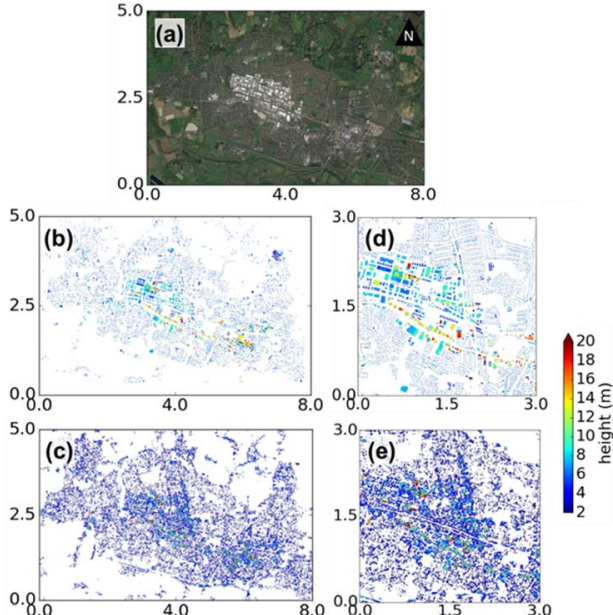
**Figure 10:** Parameters derived from the benchmark and TanDEM-X data (directly/uncorrected and corrected with suggested method, see text) for New York (mean of 1 km grid-squares from 8 sectors, Fig. 3). Map units: km. Parameter abbreviations and units: Fig. 3c. Maps for other cities are included as supplementary material.



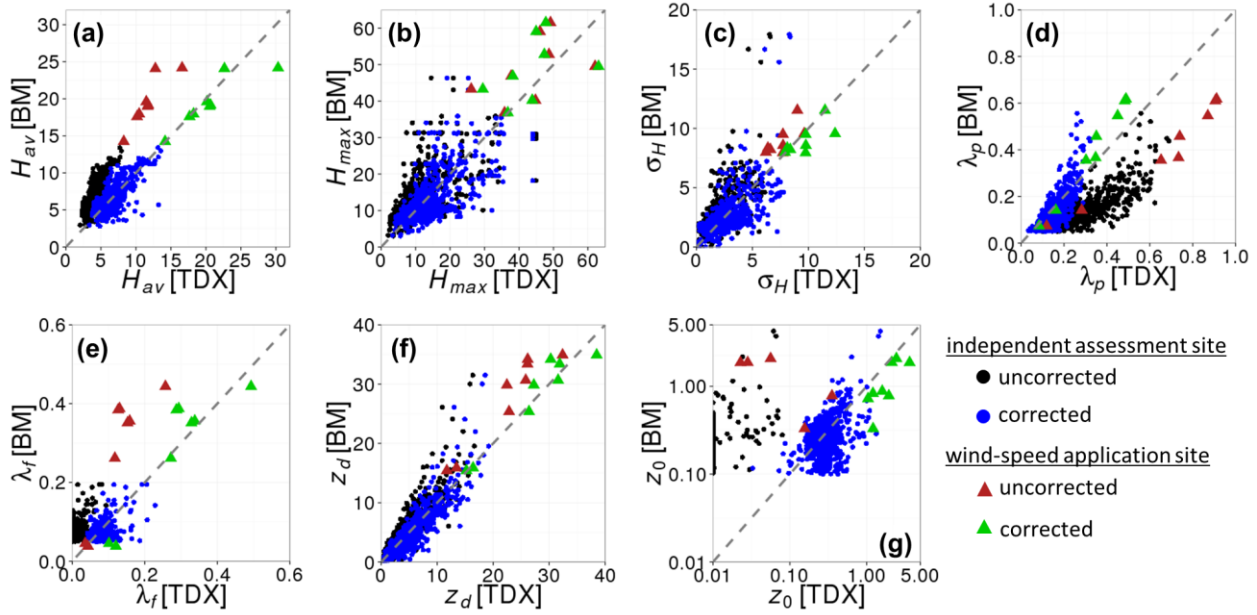
**Figure 11:** Land surface information in New York using TanDEM-X products. The water mask is used with the GUF mask (Esch et al. 2017) to indicate impervious and pervious surfaces ('settlement' and 'non-settlement' layer, respectively). For analysis of actual land cover in each mask, see Appendix A. Map units: km.

Previously, Kent et al. (2017a, 2018) used the Kormann and Meixner (2001) source area model to identify the probable upwind area and weighting for roughness parameter ( $z_d$  and  $z_0$ ) calculation. Here, the roughness parameters are also selected based on the sector of the mean wind direction for the hour (e.g. Fig. 3). This allows both the roughness parameters derived from DEMs and the impact of simplifying source area characterisation to be assessed.

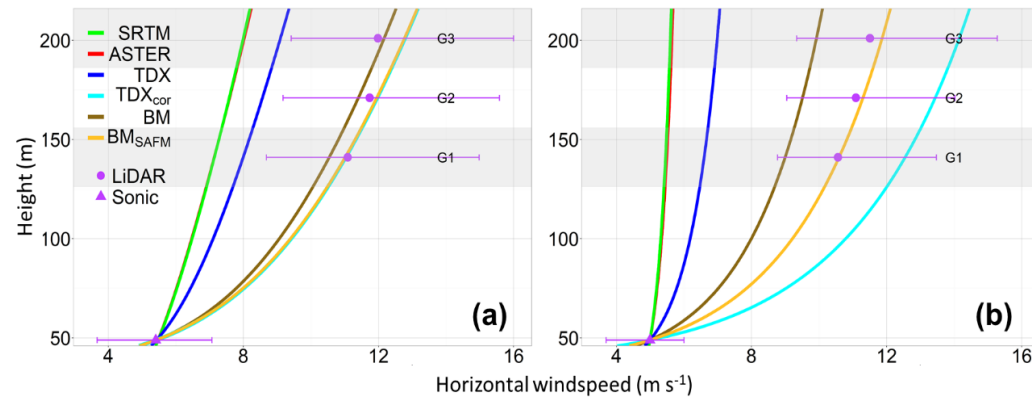
Comparing the mean wind-speed profiles (Fig. 14a) demonstrates the small roughness parameters determined from the GDEMs (Sect. 4) leads to less shear in the wind-speed profile and wind-speed underestimation. This effect is least for the TanDEM-X data given its better ability to characterise urban morphology and associated larger  $z_d$  and  $z_0$ .



**Figure 12:** Assessment site of corrections to the TanDEM-X derived parameters – Slough, UK: (a) aerial image; (b) benchmark building footprints (OS MasterMap® Topography Layer – Building Height Attribute, Ordnance Survey 2014) used to mask heights derived from the 1-m resolution EA composite lidar data (UK Environment Agency 2017); and, (c) TanDEM-X roughness-element heights within the global urban footprint mask (UFRESM) (Esch et al. 2017) from which parameters are calculated. The central 3 km x 3 km area is shown for the (d) benchmark data and (e) TanDEM-X. Map units: km. Upper left corner coordinates (WGS84, UTM30N):  $x = 661100.66$ ,  $y = 5712952.10$ .



**Figure 13:** Parameters derived from benchmark [BM] and TanDEM-X [TDX] data for (i) the separate assessment site, Slough (Fig. 12, circles here) and (ii) the wind-speed application site, central London (Fig. 6, triangles here) (Sect. 5). TanDEM-X parameters are: uncorrected (black and red symbols) and corrected (using the London-fit and multi-city procedure for  $\lambda_f$  (see text), blue and green symbols). Each point is one of eight 45° directional sectors within a 1 km grid-square – see Fig. 3 for parameter abbreviations, units and method of calculation.



**Fig. 14:** Hourly mean observed & estimated wind-speed for strong wind conditions aloft at a central London site (Fig. 6). Profiles when wind is from: (a) all directions ( $n = 245$ ), and (b) 000° – 045° ( $n = 36$ ). Observed wind speed is the average (point) and 5<sup>th</sup> and 95<sup>th</sup> percentiles (whiskers) at 49 m (sonic anemometer) and at three 30 m gates (shaded G1 – G3) by Doppler lidar. Estimated wind speed is with roughness parameters determined from different DEMs (colored lines): SRTM, ASTER, uncorrected TanDEM-X (TDX) (Fig. 13, red triangles), TanDEM-X with geometry corrected (TDX<sub>cor</sub>) (Fig. 13, green triangles), and benchmark (BM). Roughness parameters are calculated for 45° sectors with a 500-m fetch (e.g. Fig. 3a), except

for BM<sub>SAFM</sub>, which uses source area calculations. See text for more details.

Average wind-speed estimates are improved (within ~10% of benchmark data and observations) when corrected TanDEM-X roughness parameters are used (Fig. 14a, TDX<sub>cor</sub>). However, overestimates occur in the 000° – 045° sector (Fig. 14b). The TanDEM-X corrections should move wind-speed estimates towards the benchmark data (Fig. 14, BM), but a very tall feature in the TanDEM-X data in the 000° – 045° direction which is not present in the benchmark data (Fig. 6, magenta circle), leads to larger roughness parameters than the benchmark data and hence greater shear in the wind-speed profile. Historical aerial imagery does not reveal presence of a large roughness element (e.g. a crane) during the TanDEM-X data collection (2011 – 2013; tile meta-data), indicating this is possibly an artefact of the satellite derived elevation data. Although manual correction of the tall feature improves wind-speed estimates (not shown), such corrections are not plausible on a city-wide scale. The situation demonstrates that average corrected TanDEM-X data resembles the benchmark data well (Fig. 14a), but this is not always the case (Fig. 14b).

For all directions, roughness parameters derived from the benchmark data using sectors or source areas produce wind-speed estimates (Fig. 14a, BM and BM<sub>SAFM</sub>, respectively) that are within 5% of each other and the observation average. This indicates the sector-based approach may provide a reasonable simplification to source-area calculations, if forcing data or computer resources are limited (e.g. for city-scale). Again, upwind characteristics may influence this conclusion. For example, the sector-based results are poorer than the source area based results for the 000 – 045° sector (Fig. 14b) as the latter gives a larger weighting to taller buildings close to the site in this direction (Fig. 6i), which the former does not.

## 6 Discussion of GDEM comparison

The differences between the benchmark data and GDEMs can be attributed to several factors. Firstly, roughness elements or their individual parts tend to have a spatial extent which is less than the resolution of the GDEMs (especially ASTER and SRTM). Furthermore, in densely packed urban areas, layover and shadowing effects are unavoidable when using photogrammetric and interferometric techniques to retrieve surface heights. The SAR technique is also affected by foreshortening, total reflection, and multi-bounce scattering of radar in urban areas, which vary with the morphology (e.g. height and orientation) and facets of

roughness elements (Gamba et al. 2003, Stilla et al. 2003, Gamba et al. 2005, Thiele et al. 2010, Auer et al. 2011, Ferro et al. 2011, Schmitt and Stilla 2014). A combination of these factors leads to positional errors (e.g. Xu et al. 2017) with the *exact* locations and well-defined edges of roughness elements being unlikely to be resolved by the GDEM (e.g. Mercer and Gill 1998 and Figs. 6 and 7). The pixel-to-pixel comparison (Appendix B) quantifies some of these effects.

Previously, GDEMs have been demonstrated to have a density dependent signal over vegetation, with vegetation heights reported between the canopy top and bare earth surface (e.g. Hofton et al. 2006, Tachikawa et al. 2011, Li et al. 2013, Su and Guo 2014). The density dependence of the signal means the bias is expected to vary with phenology. In comparison 1, the effect of vegetation should be least for the SRTM data given it was collected during northern hemisphere winter (leaf-off, Table 1). In the other datasets, the relatively abundant, but not dense, vegetation in central London (e.g. Lindberg and Grimmond 2011a, b) may introduce a slight negative bias in the GDEM heights. In comparison 2, the GUF mask is used to retain buildings only for a consistent comparison to the benchmark data. As approximately 15% of the layer is vegetation > 2 m (Appendix A), this is another source of bias.

Temporal differences occur between collection of raw GDEM and benchmark data (cf. Table 1 and 2). The GDEM data are gathered from multiple contributing passes, with uncertainty arising; for example, with variation in surface height between passes (e.g. construction or vegetation growth) or changes in atmospheric conditions for the ASTER data (which photogrammetric techniques are sensitive to). Rapidly changing urban areas make temporal variation and differences unavoidable, which ideally benchmark data for different periods could be used to quantify. However, in comparison 1 the SRTM elevations are more similar to the benchmark data than ASTER. Given these datasets have similar horizontal resolution, but the ASTER has less temporal difference to the benchmark data, the results suggest model error outweighs the temporal effect.

## 7 Conclusions

Critical parameters for urban meteorology are derived from three global digital elevation models (GDEM): ASTER, SRTM and TanDEM-X. The TanDEM-X data are consistently most similar to central London benchmark data (20 km x 20 km area), hence the TanDEM-X data are assessed in five other cities (Auckland, Greater London, New York, Sao Paulo and Tokyo).

A moving square window extracts ground heights from the GDEMs, producing terrain models with RMSE < 4 m from benchmark data in the assessed cities. The optimum moving window width is found to be 5 x 5 pixels, except in New York, where the comparatively densely packed buildings led us to conclude that a 7 x 7-pixel window is best. Pixel-to-pixel comparisons of the digital surface models (ground and roughness-element heights) in central London found both the SRTM and TanDEM-X datasets to compare best to benchmark data (RMSE < 7 m). All the GDEMs (but especially the ASTER and SRTM) are found to increasingly underestimate the height of taller surface elements, meaning taller roughness elements (> 100 m) may be estimated with medians of up to a tenth of their true value. These conclusions hold for the TanDEM-X in the five cities, with errors largest where there is the greater proportion of densely packed tall buildings within a small plan area (e.g. New York and Tokyo).

Geometric and aerodynamic parameters are calculated for 8 directional sectors (45° width) of 1 km grid-squares, and used to evaluate the GDEMs. The average ground height is the best estimated parameter, with RMSE accuracy < 2 m across the five cities using the TanDEM-X data. The average, maximum and standard deviation of roughness-element heights and the zero-plane displacement are consistently underestimated by the GDEMs. Underestimation increases as these parameters become larger, resulting in underestimation of up to 75% using the ASTER and SRTM datasets and closer to 50% for the TanDEM-X. The apparent merging of roughness elements in the GDEMs causes the roughness-element plan and frontal areas to be over- and under-estimated, respectively. In combination, these effects produce an unreasonably small aerodynamic roughness length ( $\sim 10^{-3}$  m).

To improve the parameters derived from the TanDEM-X data, city-specific and ‘multi-city’ empirical corrections are developed. For an independent location, it is recommended to use the city-specific relations of the most similar city (e.g. for Beijing, Tokyo relations would be used), but the multi-city relation should be used to correct  $\lambda_f$ , with Eq. 2 for corrected  $\lambda_f$  values < 0.08. Across the cities, and during evaluation at a separate location, this procedure improved estimation of all geometric parameters and reduced the error in  $z_d$  by up to 20% and  $z_o$  by up to 40%. Combined with the TanDEM-X-derived water mask and GUF mask (to indicate impervious/pervious surfaces), many parameters required within urban land surface models become obtainable.

The impact of using GDEM derived roughness parameters to estimate wind speeds at up to ~10 times canopy height (from reference wind speeds observed at ~2.5 times canopy height) is assessed in central London. Results are directionally dependent, however on average, wind speeds are underestimated by up to 40% using roughness parameters from the ASTER and SRTM and 30% from the TanDEM-X. Using roughness parameters calculated with the proposed corrections to the TanDEM-X-derived geometry improves estimates to within 10% of observations. Roughness parameters derived from the benchmark data with sector-based or source area calculations led to the most accurate wind-speed estimates (within 5% of observations).

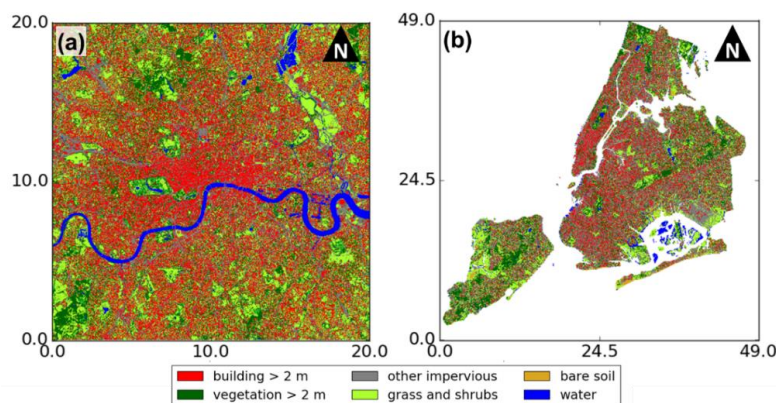
Across the GDEMs assessed, the TanDEM-X data provides the most accurate representation of urban morphology and associated wind-speed estimates. The unique morphology of different cities and resulting spatial variability of the GDEM performance means these results cannot be generalised to other cities without additional uncertainty. However, the corrections to the geometric parameters derived from the TanDEM-X model provide a basis to correct data in other cities and therefore the potential to improve the representation of urban morphology for other studies.



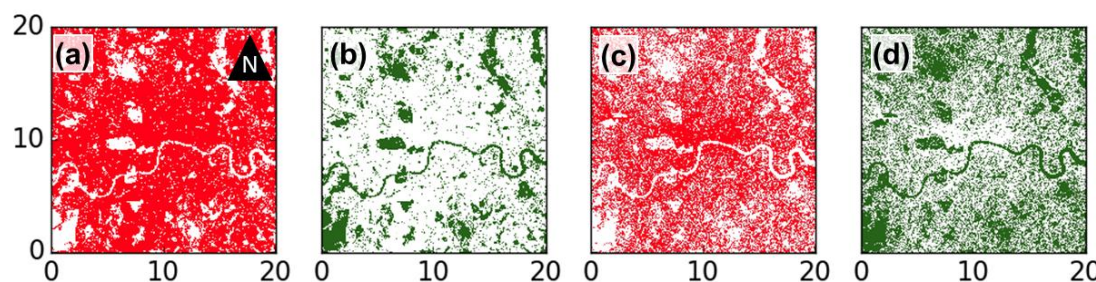
**Acknowledgements:** NERC NE/L00853X/1, Newton Fund/Met Office CSSP China funding, and TanDEM-X data (research grant: DEM\_URBAN1021, with assistance from Thomas Busche and Wieke Heldens, DLR) are gratefully acknowledged. All who assist with the London Urban Meteorological Observatory network (funding from NERC ClearfLo, EU7 emBRACE, EU7 BRIDGE, H2020 UrbanFluxes), especially: Simone Kotthaus for tower mounted observation processing, Janet Barlow and Christos Halios for provision of Doppler lidar observations. For support and provision with benchmark datasets: Danilo Mizuta (Sao Paulo Municipality), The Tokyo Metropolitan Geographic Information System Data for Urban Planning. Keith Morrison for discussion during initial analysis, suggestion of use/application for the TanDEM-X data. Omduth Coceal and Sylvia Bohnstengel for discussions regarding the work.

## Appendix A: TanDEM-X water and global urban footprint masks

Land cover information in central London and New York (Fig. A1) is used to provide insight to the land cover captured within the TanDEM-X water mask (WAM) (Wessel 2016) and global urban footprint (GUF) mask (Esch et al. 2011, 2013, 2017). The WAM accurately identifies 67% of water bodies in central London and 75% in New York, with smaller water bodies and tributaries unidentified. This result is expected, given the WAM resolves bodies of water  $> 200 \text{ m} \times 100 \text{ m}$  (Wessel 2016). The binary GUF mask designates land cover as either 'settlement' or 'non-settlement'. Visual comparison in central London (Fig. A2) demonstrates the 'settlement' layer resembles artificial surfaces, whilst the 'non-settlement' layer is more representative of vegetation, grassed areas and water bodies. Quantitative analysis reveals the settlement layer is mostly buildings and other impervious surfaces, which jointly account for 65% of the layer in London and 73% in New York (Fig. A3). The non-settlement layer is mainly impervious with 70% grassed, vegetated and small forested areas. Most buildings  $> 2 \text{ m}$  are captured by the settlement layer (less than 4% of the non-settlement layer has buildings  $> 2 \text{ m}$  in both cities), but 15% of the layer is vegetation  $> 2 \text{ m}$ .

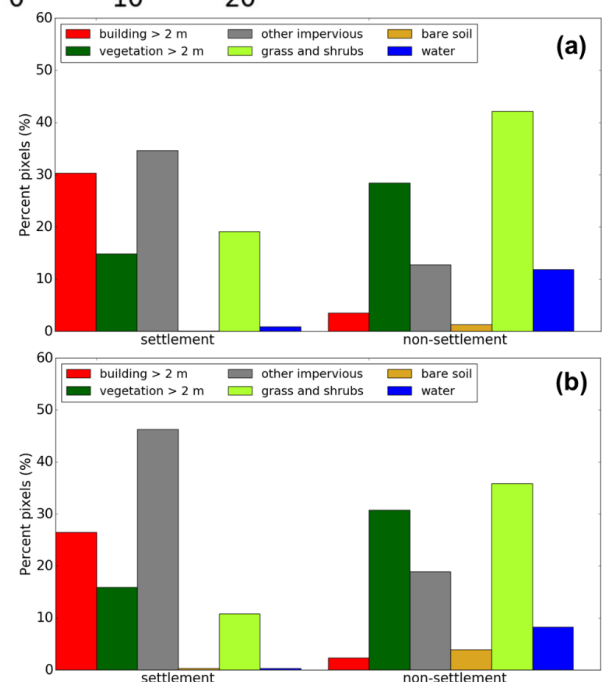


**Figure A1:** Land cover information in (a) central London and (b) New York. For London, the OS MasterMap® topography layer (Ordnance Survey, 2010) is complimented with building and vegetation height information (Lindberg and Grimmond 2011a). For New York, the New York City Landcover (2010) is used (Department of Parks and Recreation 2017), the 'tree canopy' layer is assumed to correspond to 'vegetation  $> 2 \text{ m}$ ' and 'other impervious' refers to roads, railroads and other paved surfaces. Map units: km.



**Figure A2:** (a) 'settlement' and (b) 'non-settlement' masks in the global urban footprint (GUF) mask (Esch et al. 2017) in central London. Land cover masks derived from the OS MasterMap® topography layer (Ordnance Survey 2010), comprising of: (c) built or paved surfaces and (d) tress, grasses, shrubs and water. Map units: km.

**Figure A3:** Percentage of pixels in the 'settlement' and 'non-settlement' layers of the global urban footprint (GUF) mask (Esch et al. 2017) compared to land cover information in (a) central London study area (Fig. A1a) and (b) New York (Fig. A1b). Land cover source and classes: Fig. A1. Databases resampled to 4-m pixel resolution for comparison.





## Appendix B: Pixel-to-pixel comparisons of global and benchmark DEMs

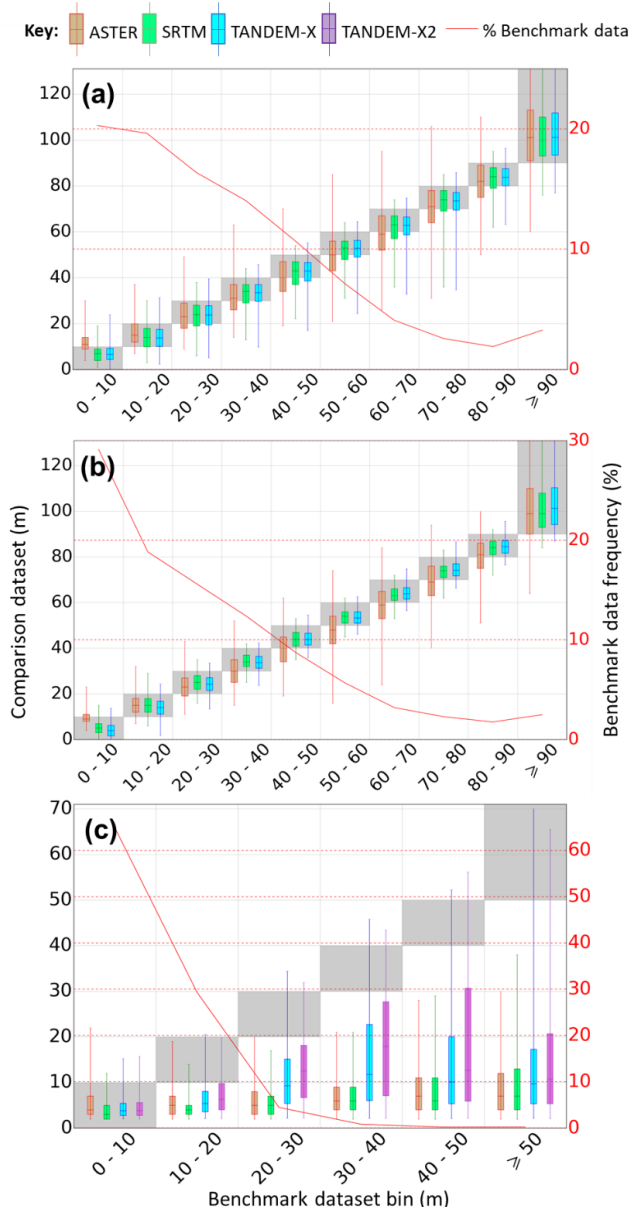
Pixel-to-pixel comparison requires a consistent pixel size, therefore all elevation models are resampled to 4-m resolution (except Tokyo, where the benchmark data only allow 5 m). Nearest neighbour resampling is used to avoid modification of originally reported heights due to interpolation (e.g. averaging). For analysis of the pixel-to-pixel comparison, the data are binned to 10 m increments. Tabulated results of the pixel-to-pixel comparison (e.g. count, percentage and error of pixels in each height bin) are provided as supplementary material.

### B.1 Comparison 1: ASTER, SRTM and TanDEM-X in London

For all GDEMs, the DSM error increases where surface heights increase (Fig. B1), which is primarily associated with the underestimation of taller roughness. However, overestimation of surface heights may also occur when comparatively coarser resolution elevations reported by the GDEMs are unable to penetrate between densely packed roughness elements (e.g. Fig. 6b-d). These over- and under-estimation effects are most obvious for the ASTER data, resulting in the DSM (and corresponding DTM) having at least 50% of the reported heights outside the range indicated by the benchmark data for each height bin (Fig. B1a, b). The differences for the TanDEM-X and SRTM are notably similar, given the horizontal resolution of the latter is over twice as coarse as the former. Both have interquartile ranges consistently within the range of benchmark DSM and DTM values.

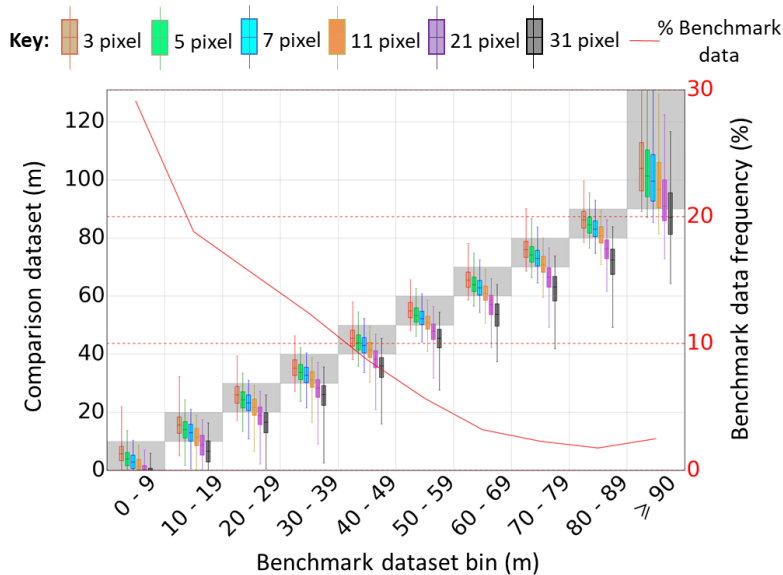
Using a 5 x 5-pixel width moving window extracts DTMs which best resemble the benchmark data for all GDEMs. The RMSE across all pixels (cf. benchmark data) is approximately 3 m for the TanDEM-X and SRTM data and 8 m for ASTER. Figure B2 shows the impact of using different sized moving windows. A smaller moving window results in consistent ground height overestimation as numerous neighboring non-ground pixels may occur within the window. Whereas, using a large window, the lowest height becomes too distant from the point of interest and is consistently lower than the ‘true’ ground height.

Median heights of roughness-elements calculated from the GDEMs range between 5 – 12 m irrespective of height bin (Fig. B1c). Most roughness-element heights in the study area are less than 20 m (95%), where the GDEMs are more accurate. This means the RMSE across all pixels is 8 m for the TanDEM-X data and 1 m larger for the other GDEMs. However, for the tallest roughness elements (> 50 m), median roughness-element heights may appear up to one tenth of their true value.



To determine if the underestimation of roughness-element heights by the GDEMs is an artefact of the DTM extraction procedure (Sect. 3.2), the ‘true’ ground heights (i.e. benchmark DTM) are subtracted from the GDEM surface models. The resulting RESMs are also compared to the benchmark data. An example with the TanDEM-X data is shown in Fig. B1c (labelled TanDEM-X2). The distribution of differences from the benchmark RESM are only slightly improved (i.e. closer to the grey boxes) when the benchmark DTM is used. The resulting RMSE across all pixels is reduced by 0.6 m, which is 7.5% of the RMSE using the extracted DTM. The underestimation of roughness-element heights therefore cannot be attributed to the DTM extraction procedure and appears inherent in the GDEMs.

**Figure B1:** Comparison between pixels in the benchmark and assessed global digital elevation models for the: (a) digital surface model (DSM), (b) digital terrain model (DTM) extracted with a 5-pixel moving window and (c) roughness-element surface model (RESM). In (c) the additional RESM assessed (TanDEM-X2) is from subtracting the benchmark DTM from the TanDEM-X DSM. For the comparisons, pixels in the benchmark dataset are divided into 10 m bins (x-axis, not inclusive of upper value labelled) and the distribution of heights in the corresponding GDEM pixels are shown for each bin (left y-axis). Whiskers are the 1<sup>st</sup> and 99<sup>th</sup> percentiles. Note, ideal distributions would be within the grey shaded areas. The percentage frequency of benchmark pixels within each bin are also shown (right y-axis, red line). Tabulated count/percentage of pixels in each height bin and errors from benchmark data are provided as supplementary material (Tables S1, S2 and S3 for the DSM, DTM and RESM, respectively).



**Figure B2:** As for Fig. B1b, but for comparison between pixels in the benchmark data and pixels of the digital terrain models (DTMs) extracted from the TanDEM-X dataset. The DTMs are extracted with Sect. 3.2 method, varying the width of the moving window by the indicated pixels (colours). Counts and error results are provided as supplementary material (Table S2c).

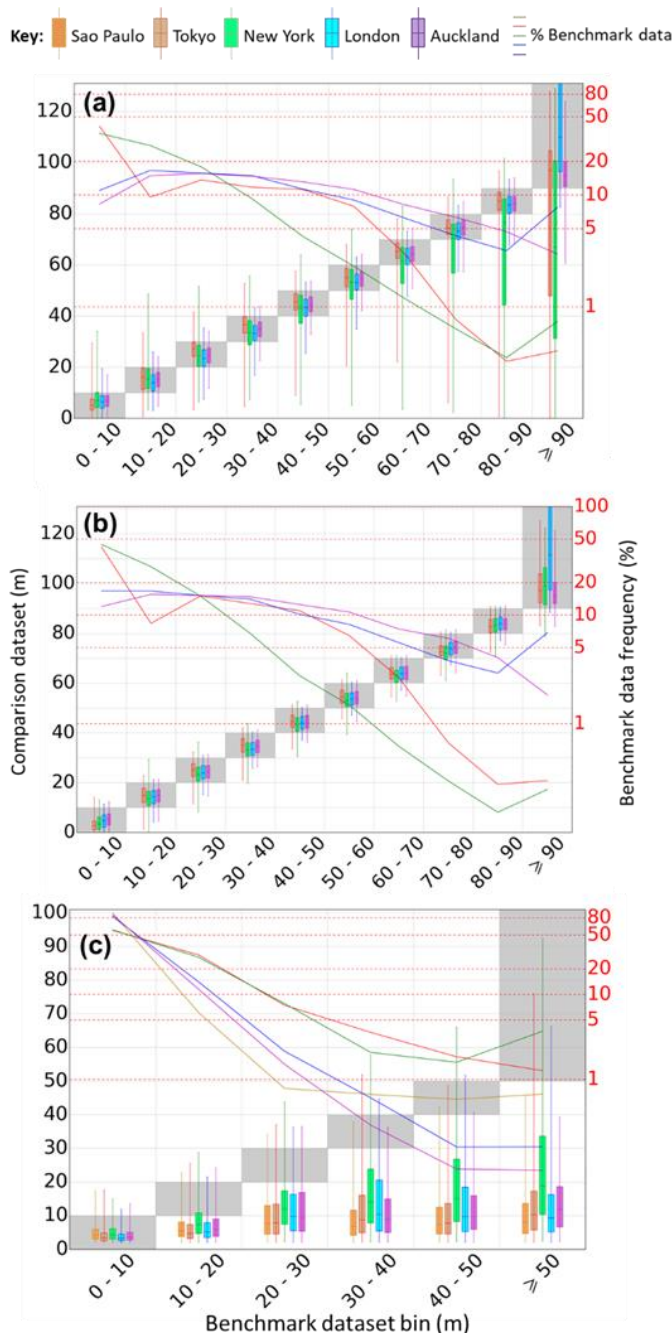
### B.2 Comparison 2: TanDEM-X in five cities

Pixel-to-pixel comparison of the TanDEM-X and benchmark data in five cities (Table 2) substantiates the findings in comparison 1. However, as more complex geometry is encountered, errors from the benchmark data become larger. The comparable morphology of London and Auckland means their respective distributions of differences from the benchmark data are similar (Fig. B3). The same is true for New York and Tokyo.

For the DSMs, the inter-quartile range of the TanDEM-X data consistently falls within the range indicated by the benchmark data, except for New York, where there is a tendency towards height underestimation (Fig. B3a) (due to underestimation of building heights). Tokyo, and especially New York, have the largest height differences from the benchmark DSM (RMSE across all pixels of 7.8 m and 11.5 m, respectively) because buildings are tall, densely packed and have small plan areas. This means the heights of surface elevations can be: (i) overestimated by greater than 20 m, due to the radar signal's inability to penetrate to ground level; and, (ii) underestimated by up to a factor of 10 where the heights of taller buildings are not resolved (see Fig. 7 and description in main text). These effects are less pronounced for the less complex morphology in Auckland and London, creating a smaller range of differences from the benchmark data (Fig. B3a) with resulting RMSE accuracies of between 4 and 5 m across all DSM pixels.

During the DTM extraction procedure, a 5 x 5-pixel moving window results in the lowest RMSE (c.f. benchmark data) for all cities except New York, where a 7 x 7-pixel window is optimum (see main text for explanation). The inter-quartile range of the resulting DTMs are within the range indicated by the benchmark data, producing RMSE across all pixels of between 2 and 4 m (Fig. B3b).

**Figure B3:** As for Figure B1, but for benchmark data in five cities compared to the TanDEM-X model with the global urban footprint mask (Esch et al. 2017) applied: (a) urban footprint digital surface model (UFDSM), (b) urban footprint digital terrain model (UFDTM) and (c) urban footprint roughness-element surface model (UFRESM). UFDTMs are extracted using Sect. 3.2 method with a 5-pixel moving window, except for New York (7 pixels). Tabulated count/percentage of pixels in each height bin and error from benchmark data provided as supplementary material (Tables S4, S5 and S6 for UFDSM, UFDTM and UFRESM, respectively).



In each city, more than half of all roughness-element heights are between 2 – 10 m (Fig. B3c). The heights of these shorter roughness elements are resolved well by the TanDEM-X data, with over 90% of elevations within the range of the benchmark values. However, roughness-element heights are increasingly underestimated as they become taller. For heights greater than 10 m, < 25% of TanDEM-X elevations tend to be within the range of the benchmark data. Beyond 50 m only a few are close to the benchmark values, with a median underestimation of up to a factor of 10. Figure 7 demonstrates the cause, which is an apparent merging of roughness elements in the TanDEM-X elevations, meaning both the height and location of the taller roughness elements are less likely to be resolved.

## Appendix C: Polynomials between TanDEM-X and benchmark data parameters

**Table C1:** Polynomial relations between parameters determined from the benchmark ( $y$ ) and the TanDEM-X ( $x$ ) datasets constrained to Eq. 1 (Sect. 3.3 method). Columns are: best fit constants ( $a$ - $c$ ), root-mean-square error for each parameter (original RMSE), following correction using the polynomial relation (corrected RMSE), normalised values (nRMSE), and data range of TanDEM-X-derived parameters ( $x$  min,  $x$  max). No correction is attempted to ground height ( $H_{av,grd}$ ). For each parameter, errors are calculated if both elevation models have  $H_{av}$  and  $H_{max} > 2$  m and the benchmark data have  $\lambda_p$  and  $\lambda_f \geq 0.05$  and  $z_d$  and  $z_0 > 0.1$ . Abbreviations and units: Fig. 3c.

Parameter	a	b	c	Original RMSE	Original nRMSE	Corrected RMSE	Corrected nRMSE	$x$ min	$x$ max
<b>(a) Sao Paulo</b>									
$H_{av,grd}$	-	-	-	-	-	-	-	-	-
$H_{av}$	0.008	-0.142	1.594	2.68	0.43	2.55	0.41	2.04	19.85
$H_{max}$	-1.83E-04	0.026	0.415	17.96	0.74	18.14	0.74	2.04	89.73
$\sigma_H$	-0.005	0.220	0.342	3.41	0.89	3.08	0.80	0.04	14.99
$\lambda_p$	-1.091	1.417	0.104	0.32	0.93	0.11	0.32	0.00	0.96
$\lambda_f$	118.780	-44.233	5.937	0.12	0.89	0.07	0.52	0.00	0.30
<b>(b) Tokyo</b>									
$H_{av,grd}$	-	-	-	1.80	0.09	-	-	-	-
$H_{av}$	0.001	-0.040	2.295	9.61	0.77	6.21	0.50	2.44	24.32
$H_{max}$	8.56E-05	-0.017	1.790	27.76	0.69	24.65	0.61	2.72	137.64
$\sigma_H$	0.002	-0.075	2.104	8.02	1.01	6.62	0.83	0.54	25.00
$\lambda_p$	0.196	-0.728	0.950	0.25	0.79	0.09	0.28	0.01	0.93
$\lambda_f$	85.820	-45.179	7.139	0.19	0.89	0.11	0.52	0.00	0.40
<b>(c) New York</b>									
$H_{av,grd}$	-	-	-	2.69	0.18	-	-	-	-
$H_{av}$	-5.40E-04	0.038	1.542	10.47	0.82	6.74	0.53	2.14	39.94
$H_{max}$	-2.30E-05	0.009	0.727	21.19	0.76	19.25	0.69	2.20	147.80
$\sigma_H$	-0.002	0.107	0.843	6.40	1.08	5.18	0.88	0.08	29.66
$\lambda_p$	0.576	-0.561	0.493	0.40	1.63	0.07	0.29	0.01	0.98
$\lambda_f$	21.199	-13.038	3.938	0.15	0.88	0.10	0.59	0.00	0.50
<b>(d) London</b>									
$H_{av,grd}$	-	-	-	1.63	0.03	-	-	-	-
$H_{av}$	1.70E-04	0.009	1.626	3.36	0.50	1.91	0.28	2.00	19.72
$H_{max}$	2.85E-04	-0.028	1.673	11.19	0.59	9.79	0.52	2.00	89.23
$\sigma_H$	0.006	-0.089	1.601	1.93	0.65	1.64	0.55	0.01	14.96
$\lambda_p$	0.920	-1.156	0.826	0.21	1.04	0.07	0.35	0.00	0.92
$\lambda_f$	151.523	-57.645	6.825	0.10	0.95	0.06	0.57	0.00	0.30
<b>(e) Auckland</b>									
$H_{av,grd}$	-	-	-	2.38	0.07	-	-	-	-
$H_{av}$	0.005	-0.078	1.530	1.84	0.32	1.22	0.21	2.02	14.51
$H_{max}$	0.002	-0.120	2.918	12.76	0.57	10.28	0.46	2.02	48.30
$\sigma_H$	0.031	-0.349	1.956	1.57	0.52	1.34	0.45	0.03	9.76
$\lambda_p$	0.088	-0.489	0.978	0.13	0.41	0.07	0.22	0.04	0.90
$\lambda_f$	725.997	-162.175	9.960	0.11	0.91	0.07	0.58	0.00	0.20
<b>(f) Multi-city</b>									
$H_{av,grd}$	-	-	-	1.95	0.06	-	-	-	-
$H_{av}$	-6.87E-04	0.057	1.099	6.29	0.74	4.76	0.56	2.00	39.94
$H_{max}$	-5.77E-06	4.91E-03	0.919	18.10	0.72	17.57	0.70	2.00	147.80
$\sigma_H$	-7.26E-04	0.049	1.120	4.58	1.03	4.04	0.91	0.01	29.66
$\lambda_p$	0.070	-0.236	0.652	0.28	1.04	0.10	0.37	0.00	0.98
$\lambda_f$	16.155	-8.884	3.135	0.14	0.97	0.11	0.76	0.00	0.50

## Appendix D: Empirical relations between benchmark data parameters

Polynomial fitting is used to provide empirical relations between the best correlated parameters in each benchmark dataset used during this work (Fig. D1). Such empirical relations may be useful when there is incomplete or unreliable information about an areas morphology. Kanda et al. (2013) describe three empirical relations between the geometric parameters of buildings within 1 km grid-squares for a 622 km<sup>2</sup> area in Tokyo (where the additional subscript  $b$  refers to buildings only):

$$\lambda_{f,b} = 1.42\lambda_{p,b}^2 + 0.4\lambda_{p,b} \quad (0.05 < \lambda_{p,b} < 0.45) \quad (D1)$$

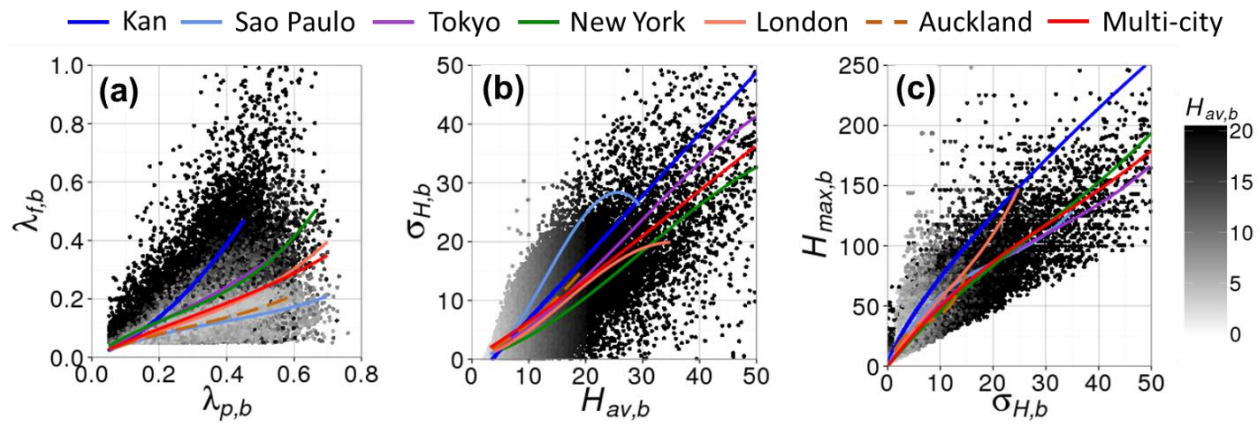
$$\sigma_{H,b} = 1.05H_{av,b} - 3.7 \quad (D2)$$

$$H_{max,b} = 12.51\sigma_{H,b}^{0.77} \quad (D3)$$

Using all benchmark data (Table 2), the same geometric parameters are found to have the best relation across the parameters considered in this work. The RMSE of polynomial fits between these parameters (Fig. D1, colored lines) in each city are less than 0.1 (Eq. D1), 4.0 m (Eq. D2) and 13.5 m (Eq. D3). The best relation is between  $H_{max,b}$  and  $\sigma_{H,b}$ , with the lowest nRMSE (between 0.3 – 0.4). See Table D1 for equation, error and data range for each fit.

Variability from the polynomial fits occurs due to the inter- and intra-city morphological variability. For example, densely-packed favela type morphology in Sao Paulo means that an increase in  $\lambda_{p,b}$  does not result in as larger an increase of  $\lambda_{f,b}$  which is found in other cities which have taller buildings (Fig. D1a). Additionally, both Fig. D1 and Kanda et al. (2013, their Fig. 2) demonstrate that as  $\lambda_{f,b}$ ,  $\sigma_{H,b}$  and  $H_{max,b}$  increase, there is increasing variability from the fitted relations. An increase in these parameters indicates more heterogeneous building morphology, suggesting use of the empirical relations becomes less certain with heterogeneity. This is supported by the largest errors from the relations being in Tokyo and New York (Table D1), the cities with the most heterogenous morphology.





**Figure D1:** Polynomial relations between geometric parameters determined from the benchmark data (buildings only, represented by subscript  $b$ ). Polynomial fits use all city-specific data for each parameter (colored lines) and the ‘multi-city’ polynomial is by Sect. 3.3 method. Auckland is dashed, as a lack of benchmark data building footprints mean the global urban footprint mask (Esch et al. 2017) is used to retain buildings only (see main text). Kanda et al. (2013) relations (Eqs. D1 – D3) are also shown (Kan). Each point is one of eight 45° directional sectors within a 1 km grid-square and is colored by the average building height ( $H_{av,b}$ ) in that 1 km grid-square. Table D1 has the equation, error and data range of fits. Parameter abbreviations and units: Fig. 3c.

**Table D1:** Polynomial relations between different geometric parameters (‘parameter’ column) determined from buildings (subscript  $b$ ) in the benchmark datasets. For each pair of parameters  $y = f(x)$ , fits are constrained to E1. 1 (Sect. 3.3 method). Given the lack of benchmark data building footprints in Auckland, the global urban footprint mask (Esch et al. 2017) is used to retain buildings only (see main text). The root-mean-square error (RMSE) and normalised RMSE (nRMSE) correspond to the fit of the polynomial function through the data points. Data range of the fit is indicated ( $x$  min,  $x$  max). Parameter abbreviations and units: Fig. 3c.

Parameter	a	b	c	RMSE	nRMSE	$x$ min	$x$ max
<b>(a) Sao Paulo</b>							
$\lambda_{t,b} = f(\lambda_{p,b})$	0.677	-0.779	0.513	0.07	0.52	0.05	0.70
$\sigma_{H,b} = f(H_{av,b})$	-0.004	0.143	-0.171	2.07	0.47	3.22	29.50
$H_{max,b} = f(\sigma_{H,b})$	0.005	-0.292	8.302	10.77	0.37	0.01	34.98
<b>(b) Tokyo</b>							
$\lambda_{t,b} = f(\lambda_{p,b})$	0.839	-0.520	0.677	0.07	0.33	0.05	0.58
$\sigma_{H,b} = f(H_{av,b})$	-2.46E-04	0.021	0.372	3.49	0.43	3.85	64.52
$H_{max,b} = f(\sigma_{H,b})$	0.0014	-0.128	6.236	13.37	0.31	0.33	64.99
<b>(c) New York</b>							
$\lambda_{t,b} = f(\lambda_{p,b})$	2.020	-1.432	0.820	0.11	0.64	0.05	0.66
$\sigma_{H,b} = f(H_{av,b})$	-2.24E-04	0.020	0.217	2.93	0.49	5.34	59.91
$H_{max,b} = f(\sigma_{H,b})$	7.73E-04	-0.064	5.145	11.28	0.37	0.08	64.58
<b>(d) London</b>							
$\lambda_{t,b} = f(\lambda_{p,b})$	0.967	-0.631	0.535	0.03	0.29	0.05	0.73
$\sigma_{H,b} = f(H_{av,b})$	-7.64E-04	0.038	0.192	1.20	0.36	3.78	80.00
$H_{max,b} = f(\sigma_{H,b})$	0.007	-0.227	7.164	7.13	0.33	1.27	67.98
<b>(e) Auckland</b>							
$\lambda_{t,b} = f(\lambda_{p,b})$	0.908	-0.827	0.525	0.03	0.25	0.07	0.69
$\sigma_{H,b} = f(H_{av,b})$	-1.94E-04	0.024	0.388	0.64	0.21	3.57	26.67
$H_{max,b} = f(\sigma_{H,b})$	0.058	-1.265	11.121	6.77	0.28	1.32	25.48
<b>(f) Multi-city</b>							
$\lambda_{t,b} = f(\lambda_{p,b})$	0.456	-0.385	0.546	0.06	0.56	0.05	0.70
$\sigma_{H,b} = f(H_{av,b})$	-3.63E-05	4.35E-03	0.601	4.19	0.96	3.22	64.52
$H_{max,b} = f(\sigma_{H,b})$	8.70E-04	-0.086	5.700	10.78	0.44	0.01	64.99

Interestingly, although Eqs. D1 – D3 were derived for an area in Tokyo, they deviate from the larger area of Tokyo considered here (Fig. D1, blue and purple lines, respectively). This demonstrates the sensitivity of empirical relations to the selected input data and exemplifies the caution which should be taken during their ‘global’ application. Fitting multi-city relations to the parameters (Fig. D1, red line) does not resolve the inter-city variability and therefore tends to have larger errors compared to city-specific fits (Table 1D).

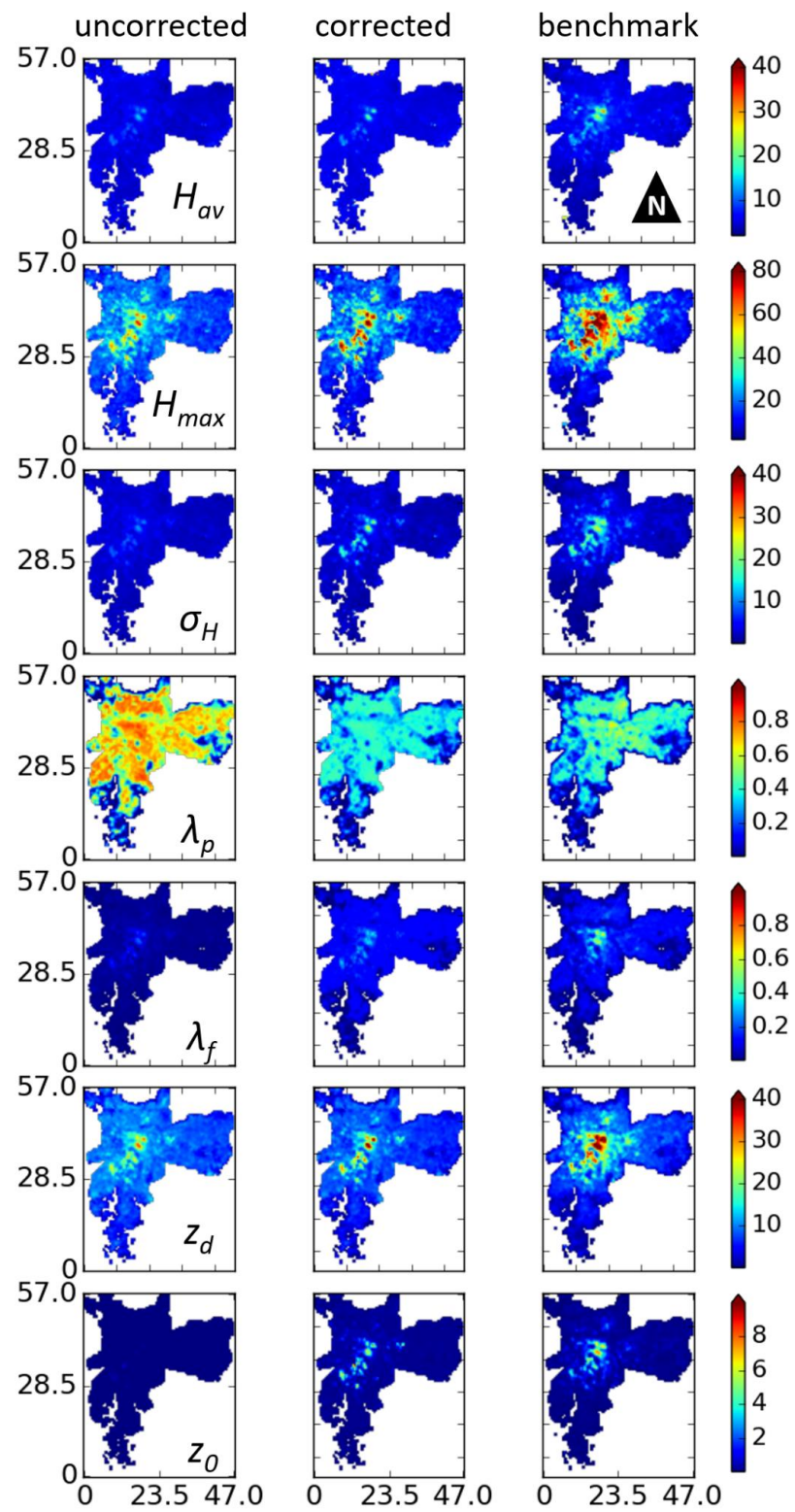
## References

- Akaike H (1974) A new look at the statistical model identification. *IEEE transactions on automatic control*. 19:716-723.
- Arnfield AJ (2003) Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *Int J Climatol* 23:1-26.
- ASTER G (2009) Validation Team: ASTER Global DEM Validation–Summary Report. METI & NASA 28.
- Auer S, Gernhardt S, Bamler R (2011) Ghost persistent scatterers related to multiple signal reflections. *IEEE Geoscience and Remote Sensing Letters*. 8:919-23.
- Bates DM, Watts DG (1988) Nonlinear regression: iterative estimation and linear approximations. In: *Nonlinear Regression Analysis and Its Applications*. Wiley, Hoboken, USA, 365 pp.
- Britter R, Hanna S (2003) Flow and dispersion in urban areas. *Annu Rev Fluid Mech* 35:469-496.
- Bürgmann R, Rosen PA, Fielding EJ (2000) Synthetic aperture radar interferometry to measure Earth’s surface topography and its deformation. *Annu Rev Earth Planet Sci* 28:169-209.
- Chen F, Kusaka H, Bornstein R, Ching J, Grimmond CSB, Grossman-Clarke S, Loridan T, Manning KW, Martilli A, Miao S, Sailor D (2011) The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. *International Journal of Climatology* 31:273-288.
- Chen Y, Su W, Li J, Sun Z (2009) Hierarchical object oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research* 43:1101-1110.
- Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association* 74:829-836.
- Crawford B, Krayenhoff ES, Cordy P (2016) The urban energy balance of a lightweight low-rise neighborhood in Andacollo, Chile. *Theo Appl Clim Oct*:1-14.



- Crawford B, Grimmond CSB, Gabey A, Marconcini M, Ward HC, Kent CW (2018) Variability of urban surface temperatures and implications for aerodynamic energy exchange in cities. *Quarterly Journal of the Royal Meteorological Society*, in press.
- CTBUH (2017) Council on Tall Buildings and Urban Habitat. Height calculator. Available: <http://www.ctbuh.org/TallBuildings/HeightStatistics/HeightCalculator/tabid/1007/language/en-GB/Default.aspx>, accessed: March 2017.
- Darmanto NS, Varquez AC, Kanda M (2017) Urban roughness parameters estimation from globally available datasets for mesoscale modeling in megacities. *Urban Climate* 21:243-261.
- Deaves D, Harris R (1978) A mathematical model of the structure of strong winds. Construction Industry Research and Information Association Report number 76, London, England.
- Eckert S, Hollands T (2010) Comparison of automatic DSM generation modules byprocessing IKONOS stereo data of an urban area. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3:162–167.
- Esch T, Heldens W, Hirner A, Keil M, Marconcini M, Roth A, Zeidler J, Dech S, Strano E (2017) Breaking new ground in mapping human settlements from space – The Global Urban Footprint. *ISPRS Journal of Photogrammetry and Remote Sensing* 134:30-42
- Esch T, Schenk A, Ullmann T, Thiel M, Roth A, Dech S (2011) Characterization of Land Cover Types in TerraSAR-X Images by Combined Analysis of Speckle Statistics and Intensity Information. *IEEE Transactions on Geoscience and Remote Sensing* 49:1911-1925
- Esch T, Marconcini M, Felbier A, Roth A, Heldens W, Huber M, Schwinger M, Taubenböck H, Müller A, Dech S (2013) Urban footprint processor—Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. *IEEE Geoscience and Remote Sensing Letters* 10:1617-1621.
- Farr TG, Kobrick M (2000) Shuttle Radar Topography Mission produces a wealth of data. *Eos, Transactions American Geophysical Union* 81:583-585.
- Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L (2007) The shuttle radar topography mission. *Rev Geophys* 45:2005RG000183.
- Fernando H (2010) Fluid dynamics of urban atmospheres in complex terrain. *Annu Rev Fluid Mech* 42:365-389.
- Ferro A, Brunner D, Bruzzone L, Lemoine G (2011) On the relationship between double bounce and the orientation of buildings in VHR SAR images. *IEEE Geoscience and Remote Sensing Letters*. 8:612-616.
- Gál T, Unger J (2009) Detection of ventilation paths using high-resolution roughness parameter mapping in a large urban area. *Build Environ* 44:198-206.
- Gamba P, Dell'Acqua F, Dasarathy BV (2005) Urban remote sensing using multiple data sets: Past, present, and future. *Information Fusion*. 6:319-26.
- Gamba P, Dell'Acqua F, Houshmand B (2003) Comparison and fusion of LIDAR and InSAR digital elevation models over urban areas. *International Journal of Remote Sensing* 24:4289-4300.
- Gamba P, Dell'Acqua F, Houshmand B (2002) SRTM data characterization in urban areas. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences* 34:55-58.
- Garuma GF (2017) Review of urban surface parameterizations for numerical climate models. *Urban Climate*, DOI: <https://doi.org/10.1016/j.uclim.2017.10.006>
- Geosampa (2017) Mapa Digital da Cidade de São Paulo – Digital Map of São Paulo City. Available: [http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/\\_SBC.aspx](http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx), accessed March 2017.
- Geiß C, Wurm M, Breunig M, Felbier A, Taubenböck H (2015) Normalization of TanDEM-X DSM data in urban environments with morphological filters. *IEEE Trans Geosci Remote Sens* 53:4348-4362.
- Goodwin NR, Coops NC, Tooke TR, Christen A, Voogt JA (2009) Characterizing urban surface cover and structure with airborne lidar technology. *Canadian Journal of Remote Sensing* 35:297-309.
- Grimmond CSB, Blackett M, Best M, Barlow J, Baik J, Belcher S, Bohnenstengel S, Calmet I, Chen F, Dandou A (2010) The international urban energy balance models comparison project: first results from phase 1. *Journal of applied meteorology and climatology* 49:1268-1292.
- Grimmond CSB, Blackett M, Best MJ, Baik JJ, Belcher SE, Beringer J, Bohnenstengel SI, Calmet I, Chen F, Coutts A, Dandou A (2011) Initial results from Phase 2 of the international urban energy balance model comparison. *International Journal of Climatology* 31:244-72.
- Grimmond C, Oke TR (1999) Aerodynamic properties of urban areas derived from analysis of surface form. *J Appl Meteorol* 38:1262-1292.
- Grimmond C, Souch C (1994) Surface description for urban climate studies: a GIS based methodology. *Geocarto Int* 9:47-59.
- Guth P. (2010) Geomorphometric comparison of ASTER GDEM and SRTM. Joint Symposium of ISPRS Technical Commission IV, Orlando, FL, USA, 10 pp.
- Hirt C (2015) Digital terrain models. *Encyclopedia of Geodesy*. Springer, Switzerland, 6 pp.
- Hofton M, Dubayah R, Blair JB, Rabine D (2006) Validation of SRTM elevations over vegetated and non-vegetated terrain using medium footprint lidar. *Photogrammetric Engineering & Remote Sensing* 72:279-285.
- Ioannidis C, Xinogalas E, Soile S (2014) Assessment of the global digital elevation models ASTER and SRTM in Greece. *Survey Review* 46:342-354.
- Jackson P (1981) On the displacement height in the logarithmic velocity profile. *J Fluid Mech* 111:15-25.
- Kanda M, Inagaki A, Miyamoto T, Gryschka M, Raasch S (2013) A new aerodynamic parametrization for real urban surfaces. *Boundary-Layer Meteorol* 148:357-377.
- Kent CW, Grimmond CSB, Barlow J, Gatey D, Kotthaus S, Lindberg F, Halios CH (2017a) Evaluation of Urban Local-Scale Aerodynamic Parameters: Implications for the Vertical Profile of Wind Speed and for Source Areas. *Boundary-Layer Meteorol* 164:183-213.
- Kent CW, Lee K, Ward HC, Hong JW, Hong J, Gatey D, Grimmond CSB (2017b). Aerodynamic Roughness Variation with Vegetation: Analysis in a Suburban Neighbourhood and a City Park. <https://doi.org/10.1007/s11252-017-0710-1>.
- Kent CW, Grimmond CSB, Gatey D, Barlow JF (2018) Assessing methods to extrapolate the vertical wind-speed profile from surface observations in a city centre during strong winds. *J Wind Eng Ind Aerodyn*. 173:100-111.
- Kormann R, Meixner FX (2001) An analytical footprint model for non-neutral stratification. *Boundary-Layer Meteorol* 99:207-224.
- Kotthaus S, Grimmond CSB (2014a) Energy exchange in a dense urban environment—Part I: Temporal variability of long-term observations in central London. *Urban Climate* 10:261-280.
- Kotthaus S, Grimmond CSB (2014b) Energy exchange in a dense urban environment—Part II: Impact of spatial heterogeneity of the surface. *Urban Climate* 10:281-307.
- Kraus K, Pfeifer N (2001) Advanced DTM generation from LIDAR data. *International Archives Of Photogrammetry Remote Sensing And Spatial Information Sciences* 34:23-30.
- Lane S, Barlow JF, Wood CR (2013) An assessment of a three-beam Doppler lidar wind profiling method for use in urban areas. *J Wind Eng Ind Aerodyn* 119:53-59.
- Li P, Shi C, Li Z, Muller J, Drummond J, Li X, Li T, Li Y, Liu J (2013) Evaluation of ASTER GDEM using GPS benchmarks and SRTM in China. *Int J Remote Sens* 34:1744-1771.
- Lindberg F, Grimmond C (2011a) Nature of vegetation and building morphology characteristics across a city: influence on shadow patterns and mean radiant temperatures in London. *Urban Ecosystems* 14:617-634.
- Lindberg F, Grimmond C (2011b) The influence of vegetation and building morphology on shadow patterns and mean radiant temperatures in urban areas: model development and evaluation. *Theoretical and applied climatology* 105:311-323.
- LINZ (2013) Land and information New Zealand. Available: <https://data.linz.govt.nz/>, accessed March 2017.
- Macdonald R, Griffiths R, Hall D (1998) An improved method for the estimation of surface roughness of obstacle arrays. *Atmospheric Environment* 32:1857-1864.
- Marconcini M, Marmanis D, Esch T, Felbier A. (2014) A novel method for building height estimation using TanDEM-X data. *Proceedings of the 2014 IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Quebec City, QC, Canada, pp. 4804-4807.
- Martilli A (2002) Numerical study of urban impact on boundary layer structure: Sensitivity to wind speed, urban morphology, and rural soil moisture. *J Appl Meteorol* 41:1247-1266.
- Mercer JB, Gill M (1998) Radar-Derived DEMs for Urban Areas. *International Archives of Photogrammetry and Remote Sensing* 32:382–388.
- Millward-Hopkins J, Tomlin A, Ma L, Ingham D, Pourkashanian M (2013) Mapping the wind resource over UK cities. *Renewable Energy* 55:202-211.
- Millward-Hopkins J, Tomlin A, Ma L, Ingham D, Pourkashanian M (2012) The predictability of above roof wind resource in the urban roughness sublayer. *Wind Energy* 15:225-243.

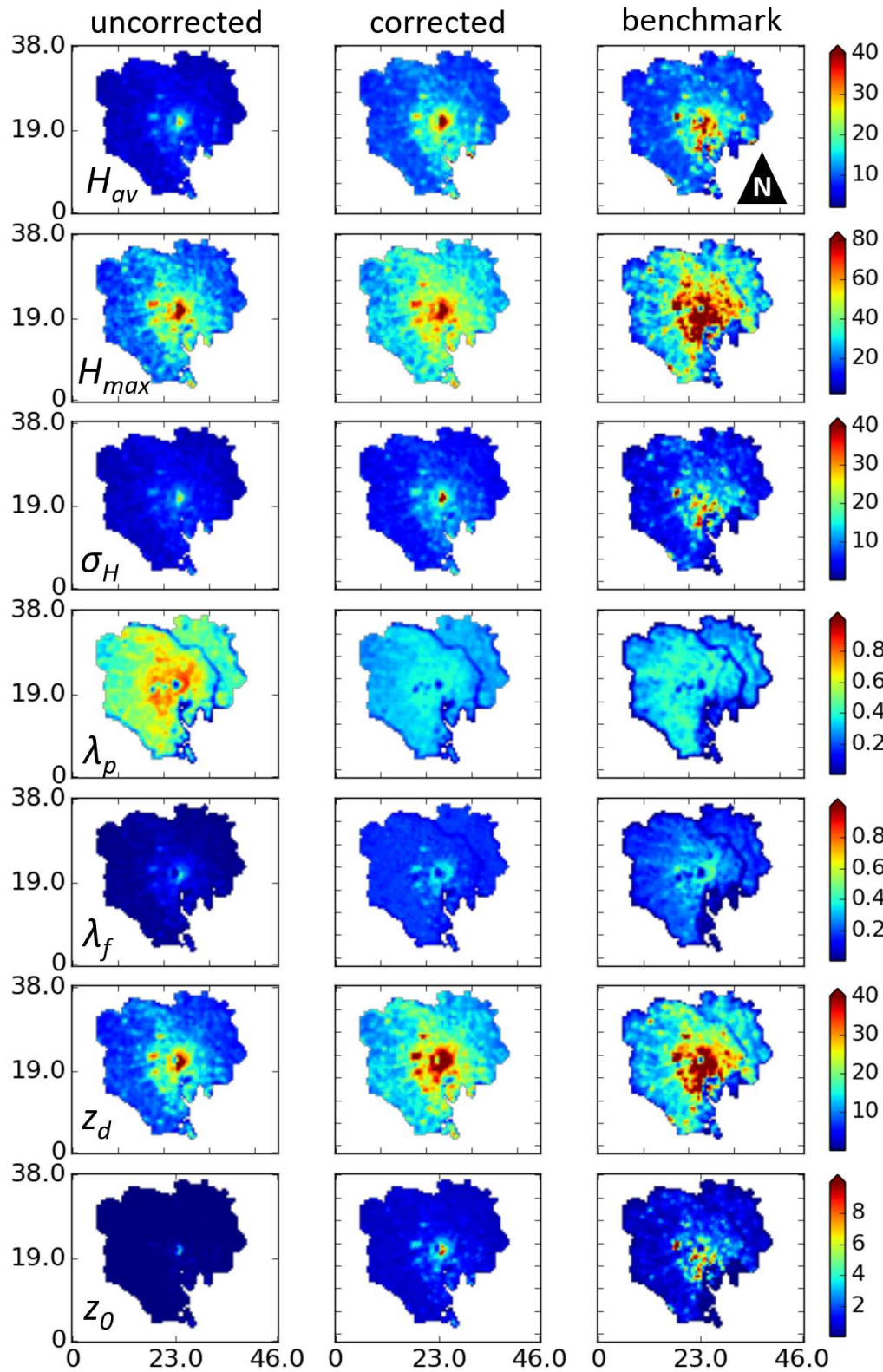
- Ng E, Yuan C, Chen L, Ren C, Fung JC (2011) Improving the wind environment in high-density cities by understanding urban morphology and surface roughness: a study in Hong Kong. *Landscape Urban Plann* 101:59-74.
- NGA (2008) National Geospatial Agency Earth Gravitational Model 2.5 minute Geoid heights. Available: [http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/egm08\\_wgs84.html](http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/egm08_wgs84.html), accessed March 2017.
- NYC DoITT (2014) New York City Department of Information Technology and Telecommunications. Available: <https://www1.nyc.gov/site/doitt/residents/gis-mapping.page>, accessed March 2017.
- Ordnance Survey (2010) MasterMap® Topography Layer, Crown Database Right 2010. An Ordnance Survey/EDINA supplied service.
- Ordnance Survey (2014) MasterMap® Topography Layer building height attribute, Crown Database Right 2014. An Ordnance Survey/EDINA supplied service.
- Rexer M, Hirt C (2016) Evaluation of intermediate TanDEM-X digital elevation data products over Tasmania using other digital elevation models and accurate heights from the Australian National Gravity Database. *Aust J Earth Sci* 63:599-609.
- Rizzoli P, Martone M, Gonzalez C, Wecklich C, Tridon DB, Bräutigam B, Bachmann M, Schulze D, Fritz T, Huber M, Wessel B (2017) Generation and performance assessment of the global TanDEM-X digital elevation model. *ISPRS Journal of Photogrammetry and Remote Sensing* 132:119-139.
- Rodriguez E, Morris CS, Belz JE (2006) A global assessment of the SRTM performance. *Photogrammetric Engineering & Remote Sensing* 72:249-260.
- Rossi C, Gernhardt S (2013) Urban DEM generation, analysis and enhancements using TanDEM-X. *ISPRS journal of photogrammetry and remote sensing* 85:120-131.
- Roth M (2000) Review of atmospheric turbulence over cities. *Q J R Meteorol Soc* 126:941-990.
- Salamanca F, Martilli A, Tewari M, Chen F (2011) A study of the urban boundary layer using different urban parameterizations and high-resolution urban canopy parameters with WRF. *Journal of Applied Meteorology and Climatology* 50:1107-28.
- Schmitt M, Stilla U (2014) Maximum-likelihood estimation for multi-aspect multi-baseline SAR interferometry of urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*. 87:68-77.
- Schreyer J, Tigges J, Lakes T, Churkina G (2014) Using airborne LiDAR and QuickBird data for modelling urban tree carbon storage and its distribution—A case study of Berlin. *Remote Sensing* 6:10636-10655.
- Small C, Sohn R (2015) Correlation scales of digital elevation models in developed coastal environments. *Remote Sensing of Environment* 159:80-85.
- Sportouche H, Tupin F, Denise L (2011) Extraction and three-dimensional reconstruction of isolated buildings in urban scenes from high-resolution optical and SAR spaceborne images. *IEEE Transactions on Geoscience and Remote Sensing* 49: 3932–3946.
- Stevens NF, Garbeil H, Mouginiis-Mark PJ (2004) NASA EOS Terra ASTER: Volcanic topographic mapping and capability. *Remote Sens Environ* 90:405-414.
- Stilla U, Soergel U, Thoennessen U (2003) Potential and limits of InSAR data for building reconstruction in built-up areas. *ISPRS Journal of Photogrammetry and Remote Sensing*. 58:113-123.
- Su Y, Guo Q (2014) A practical method for SRTM DEM correction over vegetated mountain areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 87:216-228.
- Tachikawa T, Kaku M, Iwasaki A, Gesch DB, Oimoen MJ, Zhang Z, Danielson JJ, Krieger T, Curtis B, Haase J (2011) ASTER global digital elevation model version 2-summary of validation results .
- Blackadar AK, Tennekes H (1968) Asymptotic similarity in neutral barotropic planetary boundary layers. *J Atmos Sci* 25:1015-1020.
- Thiele A, Cadario E, Schulz K, Soergel U (2010) Analysis of gable-roofed building signature in multiaspect InSAR data. *IEEE Geoscience and Remote Sensing Letters*. 7:83-87.
- Thom A (1971) Momentum absorption by vegetation. *Q J R Meteorol Soc* 97:414-428.
- USGS CMGP (2014) Unites states Geological Survey Coastal and Marine Geology Program. New York CMPG sandy 0.7m nps lidar woolpert project #73666. Available: <https://coast.noaa.gov/inventory/>, accessed March 2017.
- Varquez AC, Nakayoshi M, Kanda M (2015) The effects of highly detailed urban roughness parameters on a sea-breeze numerical simulation. *Boundary-layer meteorology* 154:449-69.
- Wang W, Xu Y, Ng E, Raasch S (2018) Evaluation of satellite-derived building height extraction by CFD simulations: A case study of neighborhood-scale ventilation in Hong Kong. *Landscape and Urban Planning* 170:90-102.
- Weekes SM, Tomlin AS (2013) Evaluation of a semi-empirical model for predicting the wind energy resource relevant to small-scale wind turbines. *Renewable energy* 50:280-288.
- Wessel B (2016) TanDEM-X Ground Segment – DEM Products Specification Document. EOC, DLR, Oberpfaffenhofen, Germany, Public Document TD-GS-PS-0021, Issue 3.1. Available: <https://tandemx-science.dlr.de/>, accessed March 2017.
- Xu Y, Ren C, Ma P, Ho J, Wang W, Lau KK, Lin H, Ng E (2017) Urban morphology detection and computation for urban climate research. *Landscape and Urban Planning* 167:212-224.
- Yan WY, Shaker A, El-Ashmawy N (2015) Urban land cover classification using airborne LiDAR data: A review. *Remote Sens Environ* 158:295-310.
- Yang L, Meng X, Zhang X (2011) SRTM DEM and its application advances. *Int J Remote Sens* 32:3875-3896.
- Zeng C, Wang J, Zhan W, Shi P, Gambles A (2014) An elevation difference model for building height extraction from stereo-image-derived DSMs. *International Journal of Remote Sensing* 35:7614–7630.



**Supplementary material to:**  
 Kent CW, Grimmond CSB, Gatey D, Hirano K (2018) Urban morphology parameters from global digital elevation models: implications for aerodynamic roughness and for wind-speed estimation. *Remote Sensing of Environment*.

**Figure S1:** As for Fig. 10, but for Sao Paulo





**Figure S2:** As for Fig. 10, but for Tokyo



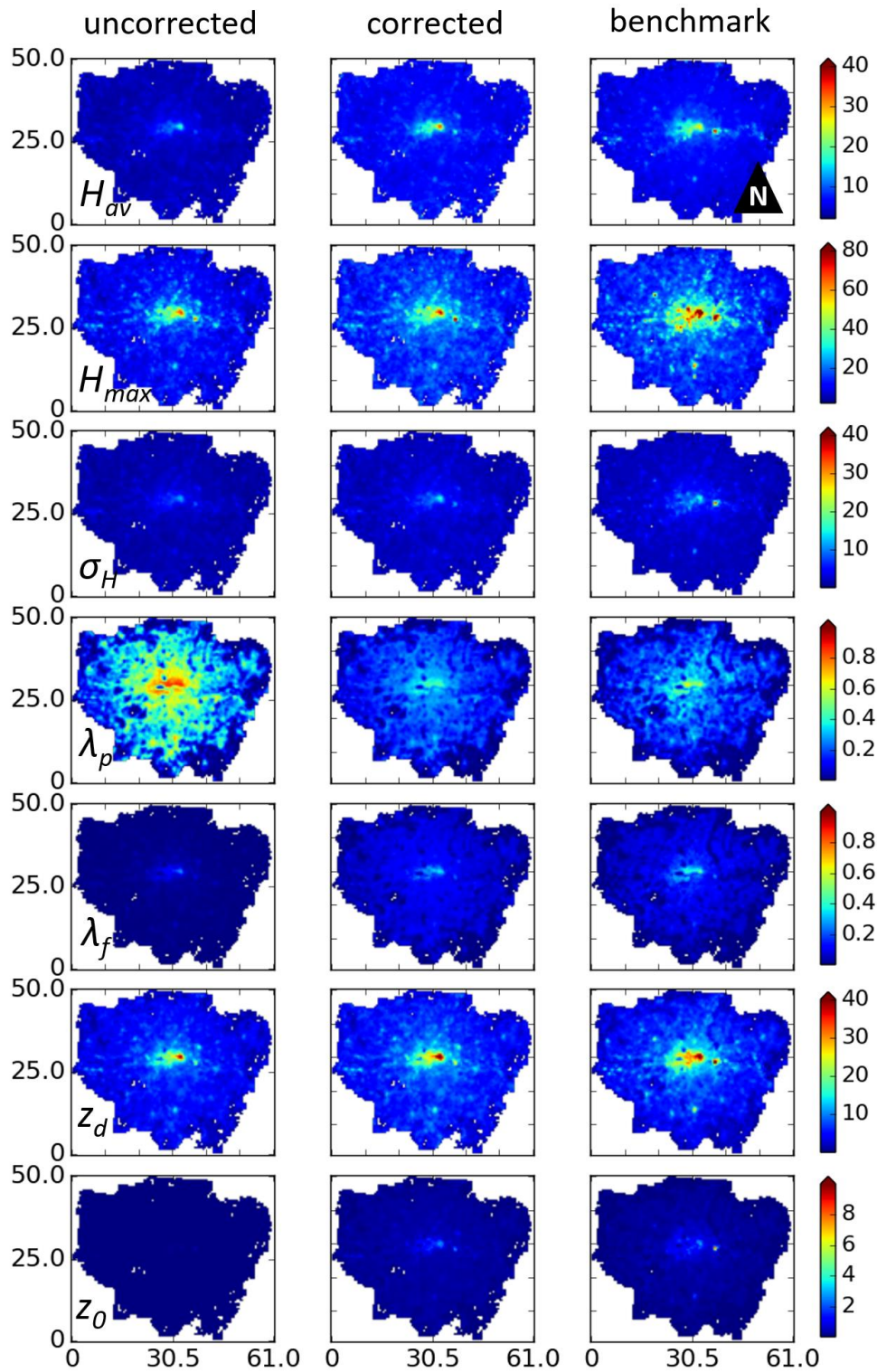
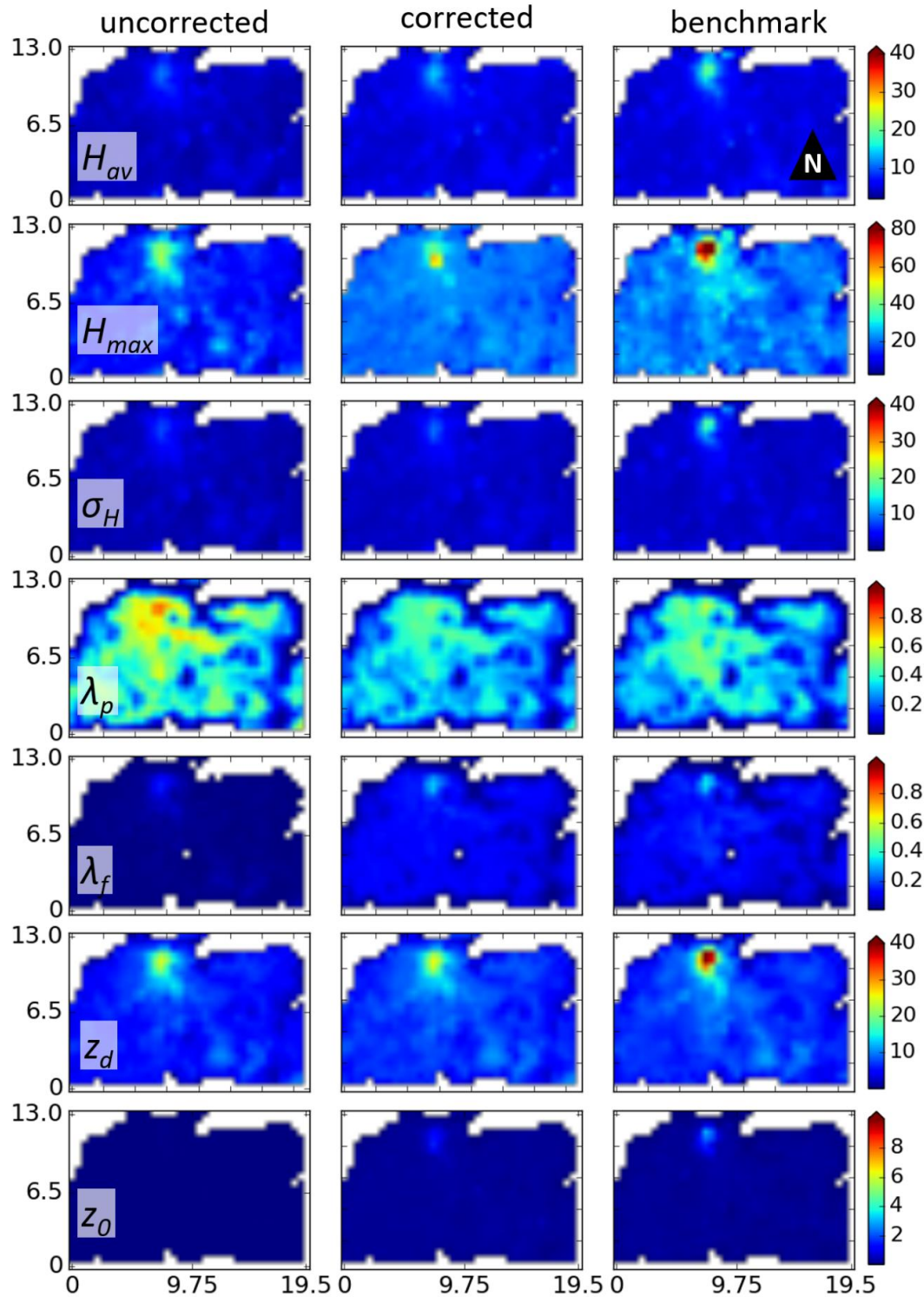


Figure S3: As for Fig. 10, but for Greater London



**Figure S4:** As for Fig. 10, but for Auckland

**Table S1:** Benchmark and global digital elevation models (GDEM) digital surface models (DSMs) compared for the 20 km x 20 km study area in central London (Fig. 4). Pixels in the benchmark dataset are divided into 10-m bins with the root-mean-square error (RMSE) from heights in the corresponding GDEM pixels shown. The pixel count is the number of pixels in the benchmark data height bin and the percentage of total pixels in each height bin is shown. The 'all' column refers to values for all pixels.

Digital Surface Model											
Height bin	0 – < 10	10 – < 20	20 – < 30	30 – < 40	40 – < 50	50 – < 60	60 – < 70	70 – < 80	80 – < 90	≥ 90	all
Pixel count	5119465	4956038	4132715	3543767	2697738	1803043	1033659	649840	480344	826751	25243360
% data	20.29	19.63	16.37	14.04	10.69	7.14	4.09	2.57	1.90	3.28	100.00
ASTER	9.12	6.59	8.31	9.98	11.15	12.54	13.14	13.29	11.69	13.95	9.79
SRTM	4.20	5.35	6.35	6.24	6.96	7.07	7.52	8.13	7.18	11.37	6.27
TanDEM-X	4.58	5.50	6.35	6.32	7.05	7.17	7.35	7.84	7.07	10.74	6.35

**Table S2:** As for Table S1, but for the digital terrain models (DTMs) extracted (Sect. 3.2 method) using the: (a) ASTER, (b) SRTM and (c) TanDEM-X models. The RMSE results for different sized moving windows are shown.

Digital Terrain Model											
Height bin	0 – < 10	10 – < 20	20 – < 30	30 – < 40	40 – < 50	50 – < 60	60 – < 70	70 – < 80	80 – < 90	≥ 90	all
Pixel count	7345642	4749633	3934730	3121852	2203549	1431317	807597	576930	442702	627475	25241427
% data	29.09	18.82	15.59	12.37	8.73	5.67	3.20	2.29	1.75	2.49	100.00
(a) ASTER											
3 pixel	8.25	6.67	7.93	8.63	9.76	11.33	11.61	10.49	9.34	8.94	8.56
5 pixel	6.33	5.09	6.88	8.54	10.14	12.27	12.55	11.35	10.02	9.16	7.90
7 pixel	5.21	4.45	6.85	9.28	11.34	13.80	14.43	13.20	11.89	10.76	8.22
11 pixel	4.05	4.37	7.80	11.21	14.07	16.92	18.54	17.38	16.25	15.04	9.75
21 pixel	3.06	5.26	10.21	15.00	19.33	22.69	26.63	27.08	26.24	25.29	13.46
31 pixel	2.73	6.00	11.78	17.49	22.86	26.69	31.85	34.41	34.40	33.92	16.22
(b) SRTM											
3 pixel	3.67	4.17	4.04	2.67	2.77	2.71	2.98	3.10	3.13	3.98	3.58
5 pixel	2.75	3.30	3.30	2.52	2.74	2.68	3.34	3.65	3.51	3.82	3.01
7 pixel	2.61	3.22	3.50	3.44	3.94	3.89	5.17	5.91	5.93	6.30	3.58
11 pixel	2.87	3.81	4.72	5.55	6.73	6.75	8.78	10.37	11.21	12.03	5.45
21 pixel	3.67	5.80	7.67	9.71	12.16	12.81	15.35	18.28	21.32	24.06	9.53
31 pixel	4.22	7.35	9.97	12.75	16.04	17.51	20.03	23.50	28.08	33.76	12.58
(c) TanDEM-X											
3 pixel	4.28	4.04	3.80	2.94	2.96	3.00	3.05	3.50	3.68	5.52	3.82
5 pixel	3.04	3.19	3.05	2.78	2.56	2.48	2.47	2.76	2.68	4.01	2.97
7 pixel	3.02	3.68	3.57	3.46	3.09	2.92	3.00	3.20	3.04	3.81	3.32
11 pixel	3.48	4.98	5.00	4.98	4.55	4.35	4.75	5.08	4.92	5.27	4.52
21 pixel	4.33	7.34	7.94	8.26	8.11	7.72	9.04	10.06	10.09	10.72	7.25
31 pixel	4.80	8.89	10.22	10.91	11.22	10.60	12.51	14.25	14.94	15.75	9.47

**Table S3:** As for Table S1, but for the roughness-element surface models.

Roughness-element surface model							
Height bin	0 – < 10	10 – < 20	20 – < 30	30 – < 40	40 – < 50	≥ 50	all
Pixel count	6930176	3136365	475974	86526	23055	24520	10676616
% data	64.90	29.38	4.46	0.81	0.22	0.23	100.00
ASTER	4.87	9.14	17.83	26.95	36.16	71.29	8.87
SRTM	3.81	9.98	18.53	27.16	36.76	70.06	9.02
TanDEM-X	3.84	8.43	14.38	21.63	32.18	63.36	7.79
TanDEM-X2	3.89	7.66	12.66	19.45	29.88	60.24	7.17

**Table S4:** As for Table S1, but for the digital surface models in global cities compared to TanDEM-X. Only pixels within the 'settlement' layer of the global urban footprint (GUF) mask (Esch et al. 2017) are compared. (Note, benchmark DSM unavailable for Sao Paulo, Table 7.2).

Digital Surface Model											
Height bin	0 – < 10	10 – < 20	20 – < 30	30 – < 40	40 – < 50	50 – < 60	60 – < 70	70 – < 80	80 – < 90	≥ 90	all
(a) Tokyo											
Pixel count	16323211	3773193	5379406	4625697	4412577	3114824	1167430	302940	127553	158040	39384871
% data	41.46	9.58	13.66	11.74	11.20	7.91	2.96	0.77	0.32	0.40	100.0
RMSE	6.43	6.00	5.82	7.07	6.86	7.24	9.62	16.61	24.17	59.05	7.83
(b) New York											
Pixel count	12485082	9698127	6250419	3234932	1485451	803604	413137	221877	122193	257828	34972650
% data	35.70	27.73	17.87	9.25	4.25	2.30	1.18	0.63	0.35	0.74	100.00
TanDEM-X	7.35	8.11	8.03	8.51	10.87	14.09	19.76	24.16	33.24	87.57	11.54
(c) London											
Pixel count	7246096	10832625	10261395	9720685	7359524	5869764	4045496	2813479	2068047	5057480	65274591
% data	11.10	16.60	15.72	14.89	11.27	8.99	6.20	4.31	3.17	7.75	100.00
TanDEM-X	3.83	4.38	4.93	4.85	5.35	5.29	5.38	5.62	5.41	6.12	4.98
(d) Auckland											
Pixel count	586681	1052977	1099178	1040001	926596	791514	576990	446253	331347	208863	7060400
% data	8.32	14.91	15.57	14.73	13.12	11.21	8.17	6.32	4.69	2.96	100.00
TanDEM-X	3.02	3.27	3.58	3.64	3.56	3.83	4.24	4.78	4.96	9.18	4.02



**Table S5:** As for Table S2, but for DTMs in global cities compared to TanDEM-X (Note, benchmark DTM unavailable for Sao Paulo, Table 7.2).

Digital Terrain Model											
Height bin	0 – < 10	10 – < 20	20 – < 30	30 – < 40	40 – < 50	50 – < 60	60 – < 70	70 – < 80	80 – < 90	≥ 90	all
<b>(a) Tokyo</b>											
Pixel count	15837951	3128953	5584790	4754432	4095289	2414506	990750	248643	102949	111643	37269906
% data	42.50	8.40	14.98	12.76	10.99	6.48	2.65	0.66	0.28	0.30	100.00
3 pixel	4.90	3.83	3.45	3.92	2.71	2.46	2.66	2.50	2.51	2.50	4.08
5 pixel	3.20	3.13	3.11	3.37	2.45	2.01	2.30	2.67	3.15	3.44	3.04
7 pixel	2.75	3.78	4.02	4.16	3.20	2.63	3.06	3.99	4.87	5.53	3.34
11 pixel	2.82	5.33	6.02	6.13	4.94	4.22	4.90	6.76	8.31	9.70	4.61
21 pixel	3.31	8.08	9.83	10.07	8.54	7.62	8.85	12.41	14.69	18.30	7.35
31 pixel	3.59	9.84	12.49	12.93	11.23	10.28	11.96	16.63	19.34	24.02	9.35
<b>(b) New York</b>											
Pixel count	15398270	9608949	5151666	2340063	951712	511258	213386	102398	52728	85342	34415772
% data	44.74	27.92	14.97	6.80	2.76	1.49	0.62	0.30	0.15	0.25	100.00
3 pixel	5.83	7.92	6.96	5.31	5.01	5.22	4.45	4.03	3.27	4.73	6.57
5 pixel	3.67	5.51	4.94	3.68	3.34	3.91	3.09	3.40	2.65	3.28	4.44
7 pixel	3.04	4.67	4.42	3.67	3.50	4.26	3.46	4.19	3.47	3.88	3.85
11 pixel	3.20	4.93	5.21	5.11	5.35	6.33	5.51	6.62	5.55	6.36	4.36
21 pixel	4.16	7.20	8.39	9.04	10.29	12.18	11.38	12.89	11.42	13.62	6.74
31 pixel	4.73	8.91	10.91	12.34	14.67	17.71	17.06	19.47	17.37	22.07	8.69
<b>(c) London</b>											
Pixel count	10817628	10802596	9868707	9105862	6626193	5338087	3628152	2476706	1903998	4471379	65039308
% data	16.63	16.61	15.17	14.00	10.19	8.21	5.58	3.81	2.93	6.87	100.00
3 pixel	2.87	2.73	2.57	1.83	1.69	1.70	1.77	1.72	1.82	1.92	2.29
5 pixel	2.16	2.30	2.34	1.92	1.75	1.69	1.69	1.67	1.68	1.92	2.04
7 pixel	2.46	2.81	2.90	2.60	2.42	2.36	2.40	2.44	2.47	2.87	2.62
11 pixel	3.15	3.92	4.16	3.90	3.78	3.75	3.99	4.11	4.24	5.06	3.92
21 pixel	4.18	5.95	6.68	6.60	6.72	6.80	7.61	7.89	8.32	10.52	6.73
31 pixel	4.76	7.41	8.63	8.77	9.21	9.39	10.64	11.07	11.82	15.44	9.07
<b>(d) Auckland</b>											
Pixel count	845505	1093274	1067701	1038192	880845	751710	525199	430462	286070	129796	7048754
% data	12.00	15.51	15.15	14.73	12.50	10.66	7.45	6.11	4.06	1.83	100.00
3 pixel	2.88	2.47	2.48	2.49	2.26	2.45	2.40	2.53	2.51	2.62	2.50
5 pixel	2.22	2.29	2.29	2.10	2.01	2.30	2.30	2.35	2.34	2.61	2.23
7 pixel	2.65	3.13	3.26	3.02	3.05	3.34	3.38	3.37	3.43	4.11	3.16
11 pixel	3.61	4.99	5.56	5.41	5.60	5.85	5.92	5.84	5.93	7.56	5.40
21 pixel	4.92	8.20	10.36	10.88	11.48	11.69	11.96	11.63	11.70	15.19	10.33
31 pixel	5.53	10.12	13.50	15.15	16.47	16.69	17.10	16.69	17.09	21.44	14.27

**Table S6:** As for Table S3, but for the roughness-element surface models in five cities compared to the TanDEM-X data only. In Sao Paulo, Tokyo, New York and London the pixels compared are 'buildings' in the benchmark data. In Auckland, the 'settlement' layer of the global urban footprint (GUF) mask (Esch et al. 2017) is the basis for comparison.

Roughness-element surface model							
Height bin	0 – < 10	10 – < 20	20 – < 30	30 – < 40	40 – < 50	≥ 50	all
<b>(a) Sao Paulo</b>							
Pixel count	17782457	1191816	152573	129783	113353	130725	19500707
% data	91.19	6.11	0.78	0.67	0.58	0.67	100.00
RMSE	3.66	7.65	16.94	27.24	35.68	54.52	7.75
<b>(b) Tokyo</b>							
Pixel count	3620592	1862005	476135	228282	117809	80664	6385487
% data	56.70	29.16	7.46	3.58	1.84	1.26	100.00
RMSE	4.20	8.34	16.09	24.71	36.58	91.51	15.34
<b>(c) New York</b>							
Pixel count	5597081	2661281	758595	202203	154448	357963	9731571
% data	57.50	27.35	7.80	2.08	1.59	3.68	100.00
RMSE	4.16	7.81	13.08	21.68	29.56	80.65	17.72
<b>(d) London</b>							
Pixel count	14111170	2392415	364963	103325	27316	27403	17026592
% data	82.88	14.05	2.14	0.61	0.16	0.16	100.00
RMSE	3.60	8.10	14.48	22.59	32.58	62.21	6.69
<b>(e) Auckland</b>							
Pixel count	3584789	477304	62302	11999	3643	3526	4143563
% data	86.51	11.52	1.50	0.29	0.09	0.09	100.00
RMSE	2.96	7.44	14.76	24.58	33.93	61.44	4.99



